University of Glasgow

Havyarimana, Enock (2023) *Understanding environmental risk factors associated with vasculitis in United Kingdom.* PhD thesis.

https://theses.gla.ac.uk/83846/

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses
https://theses.gla.ac.uk/
research-enlighten@glasgow.ac.uk

# Understanding environmental risk factors associated with vasculitis in United Kingdom

Enock Havyarimana

BSc MedSc (Hons), MPH

*Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy*

School of Infection and Immunity

College of Medical, Veterinary and Life Sciences

University of Glasgow

June 2023

# Abstract

Systemic vasculitis constitutes a range of multi-system disorders that affect small, to medium, and large blood vessels. These disorders affect 1 to 34 cases per million population each year with ANCA-Associated vasculitis (AAV) and Giant cell arteritis (GCA) being the most prevalent vasculitis in people over the age 50. The aetiology is still unknown. Recent evidence has suggested that occupation airborne exposures and serious infections may be an important risk factor for AAV. The extent to which this is true at population level is not yet clear.

The primary objectives of this thesis were (i) to investigate the long-term impact of outdoor air pollution on the onset of vasculitis, (ii) interrogate the role of geography in mediating the relationship between air pollution and vasculitis (iii) assess the temporal and seasonal variation of vasculitis onset and the possible links with environmental exposures. The method used in this thesis encompassed a systematised review, an environmental-wide association study (EWAS) approach using cross-sectional data from UK Biobank and the Scottish Morbidity Record. A series of multivariable analyses adjusted for important confounders were used to quantify the relationship between air pollution and vasculitis

Findings from the systematised review indicated that the effects of air pollution on vasculitis are variable depending on geography. It also showed that occupation airborne exposures and farming were associated 2-fold risk of vasculitis, especially for AAV. Results from UK Biobank and SMR01 suggests that long-term exposure to sulphur dioxide ($SO_2$) is associated with 6.4% and 6.9% increased odds of developing vasculitis, particularly AAV. Particulate matter ($PM_{10}$ and $PM_{2.5}$) was uniquely associated with GCA. Importantly, geography was seen to play an important role in vasculitis. Individuals from rural areas had 18% and 16% higher risk of vasculitis compared with individual from urban areas in UK Biobank and SMR01.

The temporal and seasonal analyses of AAV indicated that there are two major peaks in the incidence of AAV in Scotland. The first was seen between 1996-2000 after the introduction of antineutrophil cytoplasmic antibodies (ANCA) testing in clinical settings

in the early 1990s. The second peak was between 2017-2018 (2nd peak) and could potentially be linked with environmental agent. Overall, there was no seasonal variation in the incidence of AAV.

This thesis introduces important novel and validated results that show that outdoor air pollution may be an important risk factor of vasculitis. There is scope to build on this work in other international cohorts through data linkage of routine health data and environmental data by means EWAS study design

# Table of Contents

# List of Table

# List of Figures

# Glossary of key terms

| | |
|---|---|
| AAV | ANCA-Associated vasculitis |
| ACR | American College of Rheumatology |
| ANCA | Anti-neutrophil cytoplasm antibodies |
| AS | Ankylosing spondylitis |
| ANOVA | Analysis of Variance |
| Bz | Benzene |
| CEDA | UK National data centre for atmospheric and earth observation research |
| CHI | Community health index number |
| CO | Carbon monoxide |
| CHCC | Chapel hill consensus conference |
| DEFRA | UK Department for Environment, Food and Rural Affairs |
| EEA | European Environmental Agency |
| EGPA | Eosinophilic granulomatosis with polyangiitis |
| EMA | European Medicine Agency |
| CVD | Cardiovascular disease |
| IMD | Index of multiple deprivation |
| GCA | Giant cell arteritis |
| GPA | Granulomatosis with polyangiitis |
| LVV | Large-vessel vasculitis |
| LSOA | Layer super output area |
| MPA | Microscopic with polyangiitis |
| NIRD | Non-inflammatory rheumatic disease |
| $NO_2$ | Nitrogen dioxide |
| NOx | Nitrogen oxide |
| $O_3$ | Ozone |
| ONS | Office for national statistics |
| OR | Odds ratio |
| PAN | Polyarteritis nodosa |
| PICO | Population Intervention Comparison Outcome |
| PHS | Public health Scotland |

| | |
|---|---|
| PMR | Polymyalgia Rheumatica |
| $PM_{10}$ | Particulate matter with 10um aerodynamic diameter in size |
| $PM_{2.5}$ | Particulate matter with 2.5um aerodynamic diameter in size |
| PRISMA | Preferred Reporting Items for Systematic reviews and Meta-analysis |
| PsA | Psoriatic arthritis |
| RA | Rheumatoid Arthritis |
| SLE | Systemic Lupus Erythematosus |
| SMR01 | Scottish Morbidity Record 01 |
| SOC2000 | Standard occupation classification 2000 |
| $SO_2$ | Sulphur dioxide |
| SV | Systemic vasculitis |
| SVV | Small-vessel vasculitis |
| TAB | Temporal artery biopsy |
| TAK | Takayasu arteritis |
| UKB | UK Biobank |
| UKIVAS | UK Ireland vasculitis registry |
| US EPA | United States Environmental Protection Agency |
| WHO | World health organisation |

# Publication and conference presentations

The following publication was part of the research conducted alongside this thesis. My contribution as a leading co-author involved analysing the impact of cumulative ultraviolet B exposure and the onset of ANCA-associated vasculitis using clinical from the UKIVAS registry. The conference presentation resulted from the research conducted in thesis or as part a secondment during this PhD.

## Publications

Scott J, **Havyarimana E**, Navarro-Gallinad A, White A, Wyse J, van Geffen J, et al., The association between ambient UVB dose and ANCA-associated vasculitis relapse and onset. Arthritis Res Ther. 2022;24(1):1–14.

## Conference presentations

**Havyarimana, E**., Cullen B., Lee, D., Basu N., (2022) Long term air pollution risk factors for disease onset appear shared across systemic vasculitis and other autoimmune diseases: results from the UK Biobank, Oral presentation at the 20th International Vasculitis and ANCA Workshop, Dublin, Republic of Ireland

**Havyarimana, E**., Rutherford M., James W., Black C., Basu N., Hollick R (2022). Characterisation of systemic vasculitis outcomes across a nation: do different models, Poster presentation at the 20th International Vasculitis and ANCA Workshop, Dublin, Republic of Ireland

**Havyarimana, E**, Cullen B., Lee, D., Basu N., (2022) Air pollution exposures and ANCA-Associated vasculitis risk: A UKIVAS linkage study. Oral presentation at the Royal Society of Medicine international webinar of multi-professional faculty on vasculitis and autoinflammation diseases

# Acknowledgement

The research described in this thesis was funded by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No.813545. It was conducted using UK Biobank resource and routine data under the VOICES project. UK Biobank was established by the Wellcome Trust, Medical Research Council, UK Government Department of Health, Scottish Government and Northwest Regional Development Agency. It also received funding from the Welsh Government, British Heart Foundation and Diabetes UK. I would like to thank all the UK Biobank participants for their time and effort including UKB coordinating team. The VOICES project received funding from Versus arthritis. I would also like to thank all the vasculitis patients in Scotland for their input into this work.

Foremost, I would like to thank my supervisors (Professor Neil Basu, Dr Breda Cullen, and Professor Duncan Lee) for their continuous advice, wisdom, and support throughout my PhD. For their patience, enthusiasm, immense knowledge, and encouragement which kept me motivated. To Neil, your advice on research and my career has been priceless, you have been an incredible mentor, thank you, I've appreciated my time in Glasgow. I also would like to thank Breda Cullen and the mental and health wellbeing group for guiding me through the UK Biobank, welcoming me as one of your own and making me feel at home. My sincere thanks to Duncan Lee for your support through the many statistical component of my PhD and for propelling me to get better each time we met.

Many thanks to my colleagues at HELICAL (Albert Navarro, Alejandro Fontal, Solangez gonzalez Chiappe) and University of Glasgow (Matthew Rutherford and Flavia Sunzini) for the many meetings, writing days, and discussions which made working through the pandemic that much easier. Special thanks to Mark Little and Jessica Grene for your support in navigating all the deadlines related to funders.

Finally, I would like to thank my entire family and Gimenne. For believing in me and supporting me as always and for setting the foundation for me to pursue a PhD.

HELICAL

Funded by the Horizon 2020 Framework Programme of the European Union

# Author’s declaration

I declare that I am the sole author of this thesis, except where explicit reference is made to the contribution of others. The work in this thesis has not been submitted for any other degree at the university of Glasgow or any other institution.

Enock Havyarimana
June 2023

# Chapter 1 - Introduction

## 1.1. Overview

This chapter provides a background on the definition and classification of vasculitis. The first section provides brief history on the discovery of the different forms of vasculitis, how they are classified, named, and lastly their epidemiology and global occurrences. In closing the first section, a case is made on why the thesis focuses on ANCA-associated vasculitis and giant cell arteritis (GCA) as models for investigating environmental causes of vasculitis in the UK. A brief description on existing approaches to investigating environmental causes of adverse health outcomes is discussed. It provides the basis for the methodology that is used in the thesis. It highlights the important use of environmental-wide association studies as an important approach for investigating environmental causes of vasculitis. The second section of the introduction chapters provides a summary of the clinical definition and assessment of AAV and GCA. Their occurrence and epidemiology both in globally and in the UK including their pathogenesis. Lastly, this section ends with a detailed summary of genetic and environmental risk factors of vasculitis, with a special focus on AAV and GCA, and the gap in the literature as well as future directions. The thesis aims and research questions are provided at the end of this chapter and provides an overarching framework for this thesis and its findings.

## 1.2. Introduction to systemic vasculitis

Systemic vasculitis refers to a group of heterogenous immune mediated disorders characterised by inflammation of blood vessels, ischaemia, and necrosis of the blood vessel wall. Onset of disease is usually seen later in life, with a median age of onset being 50 years and older. Vasculitis as a disease group accounts for 1 to 4 deaths per million people per year (1),and adds enormous economic burden and pressure to society, the patient, and the wider healthcare system (2). To date, the aetiology of vasculitis is not yet clear. Current evidence suggests a complex interaction between the environment

and genes. Several risk factors have been proposed, most of which require further study. Due to its rare occurrence, systemic vasculitis aetiologies have been challenging to study, mainly because due to a lack of consistent diagnostic criteria that can be used compare studies globally and an unclear prodrome period. Regardless of these historic challenges, today, many of the vasculitis are widely studied, and new risk factors are beginning to be discovered as their prevalence grow in countries with aging population.

### 1.2.1. History

Historically, the first clinical manifestation of vasculitis was described by Robert Willan in 1808 after observing a blood spot rash appearing under the skin of one his patient(3). The rash, also known as a purpura, was described by Henoch and Schoenlein to be accompanied by a range of symptoms that included abdominal pain, pulmonary haemorrhage, arthritis, peripheral neuropathy, iritis, and nephritis (4,5).

In 1866, Kussmaul and Maier observed similar signs and symptoms in one of their patients, which they described as taking place in the perivascular sheath and outer layer of the arterial walls, and accompanied by a thickening of blood vessels, tachycardia, abdominal pain, appearance of cutaneous nodules over the trunk including general weakness caused by vasculitic neuropathy (6,7). This condition was named "periarteritis nodosa". Further research into periarteritis nodosa revealed that inflammation of the arterial wall was not fixed but was widespread across the vessel and affected the entire thickness of both the small and medium size vessels. Given this, it was appropriately renamed as "Polyarteritis nodosa" (8).

Decades later, Peter McBride (1897), described a necrotising vasculitis with granuloma lesions that affected the entire respiratory tract and glomerulonephritis (kidney) (9). Friedrich Wegener (1939) provided a clinical and pathological interpretation which was later ascribed into three distinct criteria (necrotising granulomata of the respiratory tract, generalised vasculitis,

and necrotising glomerulonephritis) and named as Wegener's granulomatosis by Churg and strauss in 1954 (10).

In 1976, DeRemee and colleagues introduced a broad classification system based on organ involvement (upper respiratory tract, lung, and kidney). The aim was to provide a more targeted description and management of the disease (11,12). In 1985, Van woude et al, described the role of autoantibodies as the main driver of disease pathology. Specifically, he highlighted that these conditions were driven by the presence of cytoplasmic autoantibodies targeting humoral immune cells (neutrophils and monocytes). These antibodies were later described as Anti-neutrophil cytoplasmic autoantibodies, associated with small vessel vasculitis (13).

For large vessel vasculitis, Hutchinson, and colleagues (1890), were the first to describe inflammation of larger sized blood vessels , specifically the temporal arteries, after observing an elderly patient with two tender red streaks with thrombosis on his scalp. They hypothesised that this inflammation was immune mediated (14) and would later described this condition as a form of thrombotic arteritis of the aged.

Decades later, Horton and colleagues (1932), coined a term Temporal arteritis after observing widespread inflammation across the artery and surrounding segments (14). In 1941, Gilmour offered the term giant cell arteritis after noting that the inflammation of the vessel was associated with increased presence of multi-nucleated giant cells in the vascular section (15). While many researchers have come up with different names to describe inflammation of temporal arteries, giant cell arteritis is the most widely used and accepted term .

### 1.2.2. Classification and definition of systemic vasculitis

The classification of systemic vasculitis has proved difficult over the past three decades, mainly due to the lack of specific and sensitive diagnostic tools to characterise the different forms of vasculitis, as well as limited knowledge about the aetiology and mechanism associated with the disease prodrome. Important distinctions must be made when referring to classification or diagnostic criteria.

Classification criteria are used to identify a homogenous group of patients for inclusion in clinical trials and other research studies while diagnostic criteria focus on listing a combination criterion or clinical findings that aim to identify with certainty the presence of a particular disease and improve its management. Many attempts have been made to classify vasculitis using our growing understanding of the disease pathology and the types of vessels involved.

To date, the most widely used system of classification and disease definition is the American College of Rheumatology (ACR) criteria, the Chapel Hill Consensus Conference on Vasculitides nomenclature (CHCC) and the European Medicine Agency four step algorithm of AAV and PAN classification criteria (16–22) (Figure 1.1). Both ACR and CHCC provide a useful guidance to rare disease physicians seeing vasculitis in its many form. The ACR is particularly useful for identifying patients for inclusion in research studies, while the CHCC has provided clarity to long standing issues related to the naming and nomenclature system of vasculitis. Both also have limitations in their application which are continually being addressed as we gain insight into the vasculitis disease pathology and aetiology.

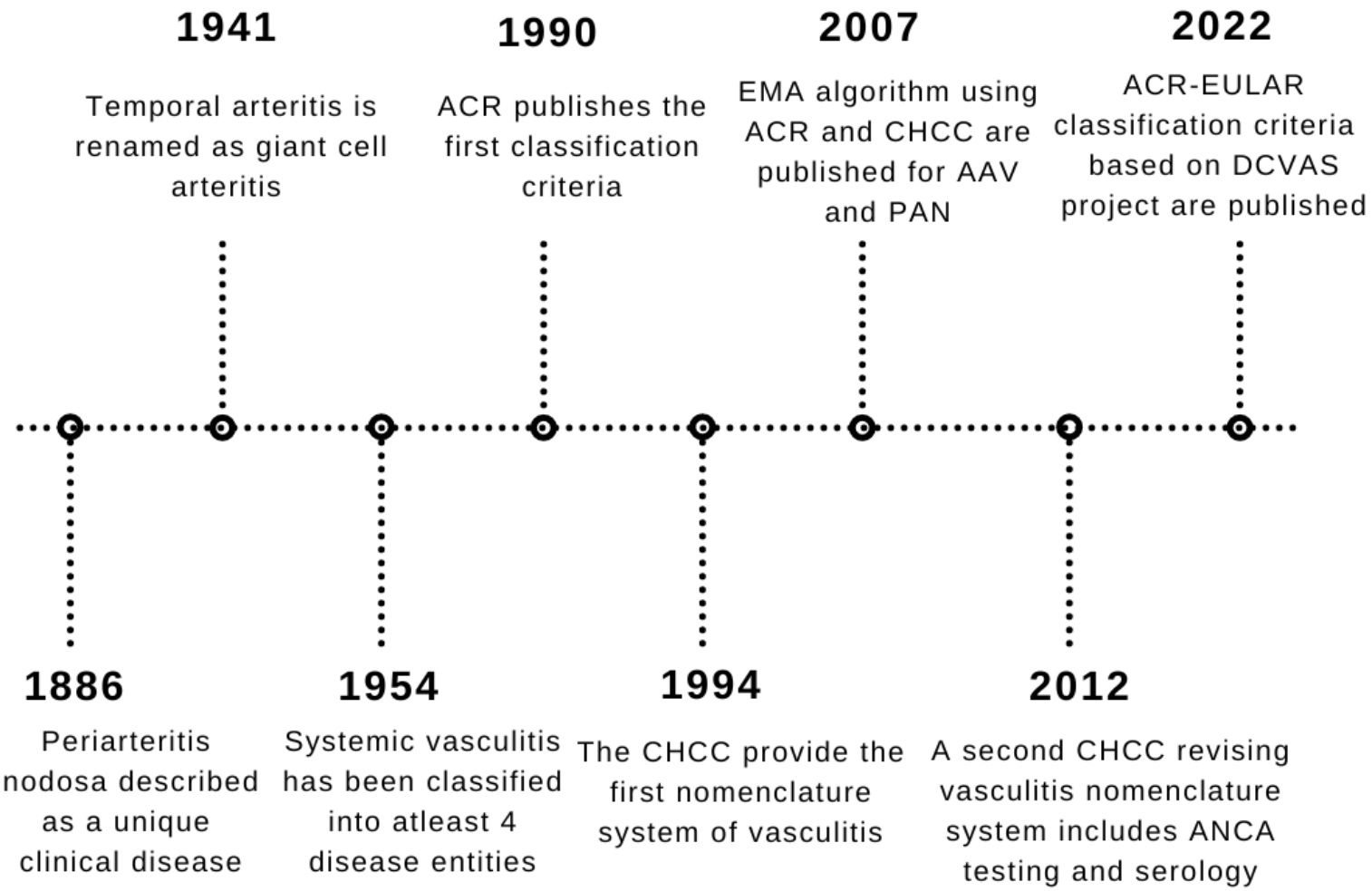



**Figure 1.1:** Timeline summarising the development of vasculitis classification criteria and nomenclature. *[The figure was created with canva.com].*

### 1.2.2.1. American College of Rheumatology (ACR,1990)

The ACR was the first classification criteria developed for vasculitis in 1990. Its aim was to improve communication between physicians regarding the different types of vasculitis and to allow for direct comparison of research studies from different region and centres(23). The ACR appointed a subcommittee tasked with generating classification criteria that was based on clinical data rather than empirical approach.

To achieve this, the committee set up a multi-centre study involving patients from 48 clinical hospital mostly from across the US. The study included seven forms of vasculitis considered clinically distinct from others: polyarteritis nodosa (PAN), Churg-Strauss syndrome (now known as eosinophilic granulomatosis with polyangiitis) , Wegener's granulomatosis (now known as granulomatosis with polyangiitis), hypersensitivity vasculitis, Henoch-Schoenlein purpura , giant cell arteritis and Takayasu arteritis (16–20). The criteria were not meant for diagnostic purposes but to capture a high proportion of patients with a particular form of vasculitis (sensitivity) from patients with other diseases (specificity).

To validate this, a study was conducted few years later where several limitations were identified and documented(24). Specifically, it was identified that the ACR criteria does not perform well for diagnosis. The criteria identified 75% of patients with vasculitis and 21% of patients with non-vasculitis. The positive predictive values for the four vasculitis studied ranged from 29% to 75% for those with a confirmed diagnosis for vasculitis(24,25).
Additionally, the ACR criteria did not highlight the role of serology and clinical imaging as surrogate markers for vasculitis definition and did not include microscopic with polyangiitis (MPA) in the disease classification. To address this, the Chapel Hill Consensus Conference on Vasculitides consensus on disease nomenclature was held in 1994 and 2012. It produced additional definitions to reflect changes in medical terminology and understanding of the disease pathophysiology.

### 1.2.2.2. Chapel Hill Consensus Conference on Vasculitis Nomenclature (CHCC 1994 and 2012)

The CHCC was first developed in 1994 and later adopted as the leading system of vasculitis nomenclature. The CHCC 1994 was assembled by a group of leading scientists and physicians from different medical disciplines with extensive experience in systemic vasculitis(26). The aim of the meeting was to reach a consensus on the names of some of the most common forms, with a particular focus on non-infectious related vasculitis. Initially, the CHCC 1994 focused on a limited number of vasculitis known at the time.

To address this limitation, a second Chapel hill consensus meeting was held in 2012. The aim was to update the disease nomenclature and to reflect changes in the naming and definitions of previously excluded vasculitis (Table 1.1) (22). In this version, there are more than 20 forms of vasculitis captured by the CHCC 2012. The definitions provide a framework for identifying and classifying these disorders, however, they do not serve as diagnostic criteria, as this requires empirical validation using observational studies that have confirmed cases of vasculitis.

**Table 1.1:** Names of systemic vasculitis listed in the chapel hill consensus nomenclature system (CHCC,2012) by affected vessels.

| Main vessel involved | Primary | Secondary |
|---|---|---|
| Large vessel vasculitis | | |
| | Takayasu arteritis | Aortitis associated with Rheumatoid arthritis (RA) |
| | Giant cell arteritis | Infection (e.g.: Syphilis) |
| Medium vessel vasculitis | | |
| | Polyarteritis nodosa | Hepatitis B virus-associated vasculitis |
| | Kawasaki disease | |
| Small vessel vasculitis | | |
| | ANCA-associated vasculitis | Vasculitis secondary to RA, SLE, Sjogren syndrome |
| | Microscopic polyangiitis | Drug-associated immune complex vasculitis |
| | Granulomatosis with polyangiitis | Drug-associated ANCA-associated vasculitis |
| | Eosinophilic granulomatosis with polyangiitis | Cancer-associated vasculitis |
| | Immune complex small vessel vasculitis | |
| | Anti-glomerular basement membrane disease | |
| | Cryoglobulinaemic vasculitis | |
| | IgA vasculitis (Henoch - Schoenlein purpura) | |
| | Hypocomplementemic urticarial vasculitis (anti-C1q vasculitis) | |
| Variable vessel vasculitis | Behcet's disease (BD) | |
| | Cogan's syndrome | |
| Single organ vasculitis | Cutaneous leukocytoclastic angiitis | |
| | Cutaneous arteritis | |
| | Primary central nervous system vasculitis isolated aortitis | |

### 1.2.2.3. European Medicine Agency algorithm (EMA, 2007)

The EMA algorithm was developed and introduced in 2007 following the recognition that the application of ACR and CHCC was proving difficult when interpreting and comparing epidemiological data for ANCA-Associated vasculitis (AAV) and polyarteritis nodosa (PAN) (27). For example, ACR has shown to have a high number of patients being classified into more than one disease while CHCC 1994 has many patients that remain unclassified when applied in isolation. To address these limitations, a group of physicians interested in the epidemiology of vasculitis met at the European Medicines Agency in September 2004 and January 2006. The main objective was to develop a consensus method and an algorithm that would use both the ACR and CHCC criteria to allow for adequate comparison of clinical data from epidemiological studies not confounded by the type of classification and nomenclature system(27). From this meeting a consensus on four-step algorithm merging the ACR 1990 and CHCC 1994 definition was constructed.

The algorithm was applied on patients with a diagnosis for small or medium vessel vasculitis and no other conditions. This was to minimise misclassification and sensitivity of algorithm. The patient had at least been followed for 3 months where possible and the algorithm and had clinical features that were present during the course of disease. Furthermore, the algorithm includes surrogate markers of ANCA vasculitis and was successfully validated on 99 patients with AAV/PAN who were appropriately classified into a single diagnosis. Today, the EMA algorithm is widely used in epidemiological studies and is reproducible and has fewer patient cases unclassified and with no overlapping diagnoses for AAV and PAN(28,29).

### 1.2.2.4. Diagnostic and Classification Criteria for Vasculitis (DCVAS) study

The DCVAS study was conceived in 2013 after recognising that in the absence of a validated criteria for vasculitis, has meant that the effectiveness of CHCC and ACR classification criteria to achieve diagnostic accuracy and improve early detection and management of patients living with vasculitis may be limited(30). To address these limitation, DCVAS a multinational observation study developed to validate the diagnostic and classification criteria of primary vasculitis aims to address existing limitation in nomenclature and classification criteria and guide future diagnostic criteria for vasculitis. Specifically, the study uses existing approach of vessel size as method of classification, incorporates detailed clinical data and the latest diagnostic tools like ANCA serology testing, biopsy, and imaging. Importantly, DCVAS uses the American College of Rheumatology and the European League against Rheumatism (EULAR) guideline on classification criteria development and has recently helped inform the 2022 ACR/EULAR classification criteria for AAV (31–33). These classification criteria were informed by clinical data from more than hundred centres across 31 countries in Asia, Europe, North America, Oceania, and South America, making it one of the more generalisable criteria capturing a wide array of disease presentation and occurrence of vasculitis in different populations across globe.

## 1.2.3. Epidemiology of systemic vasculitis

### 1.2.3.1. Time trend

The incidence of vasculitis has grown over the past four decades. This growth can be explained by improvements in disease nomenclature and increased recognition by clinician about the vasculitis clinical features and pathology. These changes have also improved overall case ascertainment of both old and existing cases. This improvement in case detection has also led to higher prevalence of some of the common small and large vessel vasculitis.

Additionally, the growth in number of vasculitis patients have increased the demand for funding to integrate national registries for multiple rare disease research. One such initiative is FAIRVASC, a semantic web platform that aims to integrate rare disease registry, with AAV being used as a case study(34). These initiatives are important for improving the overall surveillance of vasculitis, and addresses important questions related to aetiology, pathology and outcomes associated with vasculitis(30,35).

While there are many benefits associated with a rise in number of registries containing patients with vasculitis, vasculitis research faces several challenges related to the timing and design of its epidemiology. One such challenge is the collection of epidemiological data with sufficient sample size. Collecting data on a rare disease like vasculitis takes a long time and requires careful planning. Such ambition is often challenged by limited investment in resources to support mass surveillance of rare diseases, as often the case in low-middle income countries. A lack of diagnostic and standardised tools and criteria for data collection and disease nomenclature make it difficult to compare incidence data from different regions.

Regardless of these challenges, two important approaches are commonly used: the first, is the use of case cohort studies to recruit and follow patients

from at risk population (e.g.: those 50 years and older) and geographical-defined areas. These patients are often recruited from a single or several specialist centres covering a diverse population in terms of age, sex, income, population density. The second, is the use of administrative health data that includes hospital or primary care data and data from national insurance database with whole population coverage. Both approaches have advantages and disadvantages and are discussed below

### 1.2.3.2. Approaches to incidence measures and design.

#### 1.2.3.2.1. Case Cohort

Case cohort studies are commonly used in vasculitis to characterise regional and country-level incidence of the different forms of vasculitis over a given period. With this design, patients are prospectively or retrospectively followed and are often reported in aggregate form to protect the privacy of patients. Patients are most often sourced from regional hospital registries or biobanks, where consent has been provided.

The benefits of using this approach is that it allows for a detailed study of geographic and temporal clusters in disease incidence, including prognostic outcomes (like multimorbidity and mortality) and effectiveness of treatment (36,37). Large biobanks are a great tool for identifying cases and groups and populations at particular risk. Example of such resources are the Swedish Biobanks (38), UK Biobank (39), the China kadoorie biobank (40). Biobanks when combined with national patient registers and census records provide a rich resource where major health determinants associated with poor health can be studies including their associated mechanisms (41,42).

Important limitations facing case cohort studies are low sample size and limited ability to achieve statistical power need when study new risk factors associated with poor health. Patient from geographically dispersed populations with insufficient clinical service are also underrepresented in cohort studies.

Additionally, patients living with vasculitis experience high rates of multimorbidity that can serve as potentially mediators when modelling new aetiology associated with disease onset. Patient enrolment is less supervised in some registries; follow up requires continual investment in human and financial capital and may be difficult to motivate for when working with such rare diseases. Limited support in this regard can reduce patient retention and attrition, thereby propagating missing or incomplete data that may serve as a source of bias (36).

#### 1.2.3.2.2. Administrative health record

Administrative data collected for clinical, financial, and healthcare management have shown great utility for public health research (39). This data can be linked to demographic, social, environmental, and other public datasets that are geocoded. Linking these datasets for rare disease research has many benefits. They can provide substantial insight into the social and environmental determinants of health. Importantly, they provide a sample size representative of the larger population. There is better retention of patients, and the spatial coverage is much wider which is essential for rare diseases surveillance. Administrative health data can provide detailed evaluation of temporal variation in the incidence and prevalence of vasculitis including potential aetiology of disease. This data source is great for accessing of hard-to-reach populations like those living in remote and rural areas and can allow cost effective research.

The challenge lies in its limited ability to allow for diagnostic specificity due to methods used for record keeping. For example, most outcome data recorded at hospitals are coded and standardised in an international classification of disease (ICD) format. Such coding limits diagnostic accuracy essential for reporting the incidence of different forms of vasculitis. Missing data can also limit further study into outcomes associated with disease. This is especially true for patients who fail to interact with a service because of diagnostic delay and limited access to specialised services (43) often as the case in rural populations and in countries with limited resources.

### 1.2.3.3. Incidence and prevalence

Table 1.2. provides a summary of the annual incidence of primary vasculitis using studies published across the globe. A main observation can be seen; the overall incidence of vasculitis appears to vary by geography, latitude, and genetic ancestry. The most prevalent form of vasculitis is giant cell arteritis (GCA) and ANCA- Associated vasculitis (AAV). Both are common in populations of Northern European Ancestry with the median age of onset being between the age of 50 to 60. The incidence of giant cell arteritis starts from 0.7 to 4.0 cases per million population, with higher incidence being reported in Scandinavia and lower incidence in Italy and South Australia(44).

Similarly, the incidence of ANCA-associated vasculitis seen in northern European countries like Germany, Norway, and with the US account for the highest number of new cases globally. This ranges from 2.1 to 34.0 cases/million population for GPA, 2.7 to 18.2 cases/million population for MPA and 0.64 to 4.0 cases/million population for EGPA. EGPA is the least common form of AAV.

The other more prevalent forms of vasculitis are Kawasaki disease, Takayasu arteritis and Anti-glomerular basement membrane disease. Kawasaki disease is prevalent in children with 85% of those affected being under the age of 5 years old (45). Takayasu arteritis (TA) is seen in adults (30- 40-years old) (20, 21). Both conditions have a high prevalence in populations of Southeast Asian Ancestry where they were initially discovered (46,47). The global incidence of Kawasaki disease starts ranges from 0.5 to 26.5 cases per million population. Japan (24.3 and 26.4 cases per million in 2011 and 2012) and South Korea (13.4 cases per million in 2012) have recorded the highest incidence annually compared with the US (0.2 case per million) and Europe (1 case per million) (48).Interestingly, the overall incidence of Takayasu arteritis ranges from 0.4 to 3.4 cases per million population, with recent data highlighting a rise in incidence TA in Europe, particularly Turkey (49).

It is worth noting that there are only few studies reporting the incidence of Takayasu arteritis globally. This is due to long diagnostic delays and limited recognition by physician operating in non-centralised healthcare systems, for example Asia and South America. Similar challenges are true for other rare vasculitis type like variable vessel vasculitis and single organ vasculitis whose incidence is much lower and very rare compared with most small and large vessel vasculitis.

For instance, studies reporting on the incidence of IgA vasculitis, Cryoglobulinaemic vasculitis, Cogan's syndrome and Cutaneous leukocytoclastic angiitis report an overall incidence of 0.1 to 1 case per million population annually (44). To date, there are no data from regions such as Oceania, South America, middle east, and Sub-Sahara Africa. Their rare occurrence means it is difficult to study their aetiology.

**Table 1.2:** The global incidence of primary vasculitis

| Main vessel involved | Primary | Age of onset (years) | Annual Incidence/ million population | References |
|---|---|---|---|---|
| Large vessel vasculitis | | | | |
| | Takayasu arteritis | 30-40 | 0.4 to 3.4 | (50) |
| | Giant cell arteritis | 50-60 | 0.7 to 14.0 | (44,51) |
| Medium vessel vasculitis | | | | |
| | Polyarteritis nodosa | 50-60 | 0.5 to 0.9 | (52) |
| | Kawasaki disease | <5-8 | 0.5 to 26.5 | (48,53) |
| Small vessel vasculitis | | | | |
| | ANCA-associated vasculitis | 40 - 65 | 24.7 to 33.0 | (51,54) |
| | Microscopic polyangiitis | 50-65 | 2.7 to 18.2 | (51,55,56) |
| | Granulomatosis with polyangiitis | 50-60 | 2.1 to 34.0 | (51,57,58) |
| | Eosinophilic granulomatosis with polyangiitis | 40-50 | 0.1464 to 4.0 | (51,54,58) |
| | Immune complex small vessel vasculitis | ND | ND | ND |
| | Antiglomerular basement membrane disease | 60-70 | 0.5 to 2.0 | (59) |
| | Cryoglobulinaemic vasculitis | 45 -65 | 0.4 to 1.0 | (60,61) |
| | IgA vasculitis (Henoch - Schoenlein purpura) | 10 - 17 | 0.2 to 0.7 | (62,63) |
| | Hypocomplementemic urticarial vasculitis (anti-C1q vasculitis) | 50 | 0.1 to 0.7 | (64) |
| Variable vessel vasculitis | Behcet's disease (BD) | 30 | 0.1 to 0.4 | (65,66) |
| | Cogan's syndrome | ND | ND | ND |
| Single organ vasculitis | Cutaneous leukocytoclastic angiitis | 45 | 0.1 to 0.4 | (60) |
| | Cutaneous arteritis | ND | ND | ND |
| | Primary central nervous system vasculitis isolated aortitis | ND | ND | ND |

**ND – No definite data available

## 1.3. ANCA-associated vasculitis and Giant cell arteritis as models for investigating environmental causes of vasculitis.

### 1.3.1. The case for ANCA-associated vasculitis and Giant cell arteritis

It is widely accepted that the aetiology of vasculitis is a result of the dysfunctional interplay between a person's genes, their epigenetics, and environment (67,68). This interaction between the environment and genes can be assessed using a retrospective observational study focused on population, their environment and biology. For vasculitis, AAV and GCA provide such opportunity because of their large prevalence compared with other vasculitis (51). Their sample size has been growing over the past four decades and many health registries and administrative health records have started to report their incidence, while addressing the challenges of disease classification (69–71). This is particularly true for regions with more structured health system and record keeping including an ageing population (e.g.: Europe and Japan) (Figure 1.2.). Furthermore, AAV and GCA are the most studied vasculitis with much greater evidence about their prodrome and aetiology, and including biological mechanisms associated with the onset of disease. One such example important for the study of vasculitis and its aetiology is the work published from Sweden by Berglin et al.,2021 and Johansson et al.,2022 (28,72). In these studies, the authors harnessed the power of five Swedish biobanks by linking them with the national patient register. The aim was to characterise the preclinical-prodromal changes in biomarkers of AAV until the initial manifestation of disease and diagnosis. Results from these studies revealed the onset of AAV is predated by significant increases in antineutrophil cytoplasmic antibodies (ANCA) and Complement effector 5 (C5) levels that range from 3 months to 7.7 years and 5 months to 5.5 years before the onset of first symptom, respectively (28,72). Such evidence highlights the role of the gene-environment interaction as initiating this transition from a quiescent phase of disease to an active state of disease. Similarly, this hypothesis can be applied

to GCA where environmental exposures have shown significant association with surrogate markers of disease. For example, an increase in erythrocyte sedimentation rate (ESR), interleukin-6 (IL-6) and C-reactive protein (CRP), the three important surrogate markers of GCA disease activity(73), have shown to be associated with air pollution, both gaseous and particulate matter (74,75).

This interaction between the environment and disease biology needs further study. There is some evidence to support that the onset of AAV is associated with occupational airborne particles and dusts released after mass explosions (76–78). The extent to which this is true at a population level is not yet clear. By harnessing the power of the growing registries in UK, this thesis will address these gaps and further describe the role of population airborne exposure in the onset vasculitis.

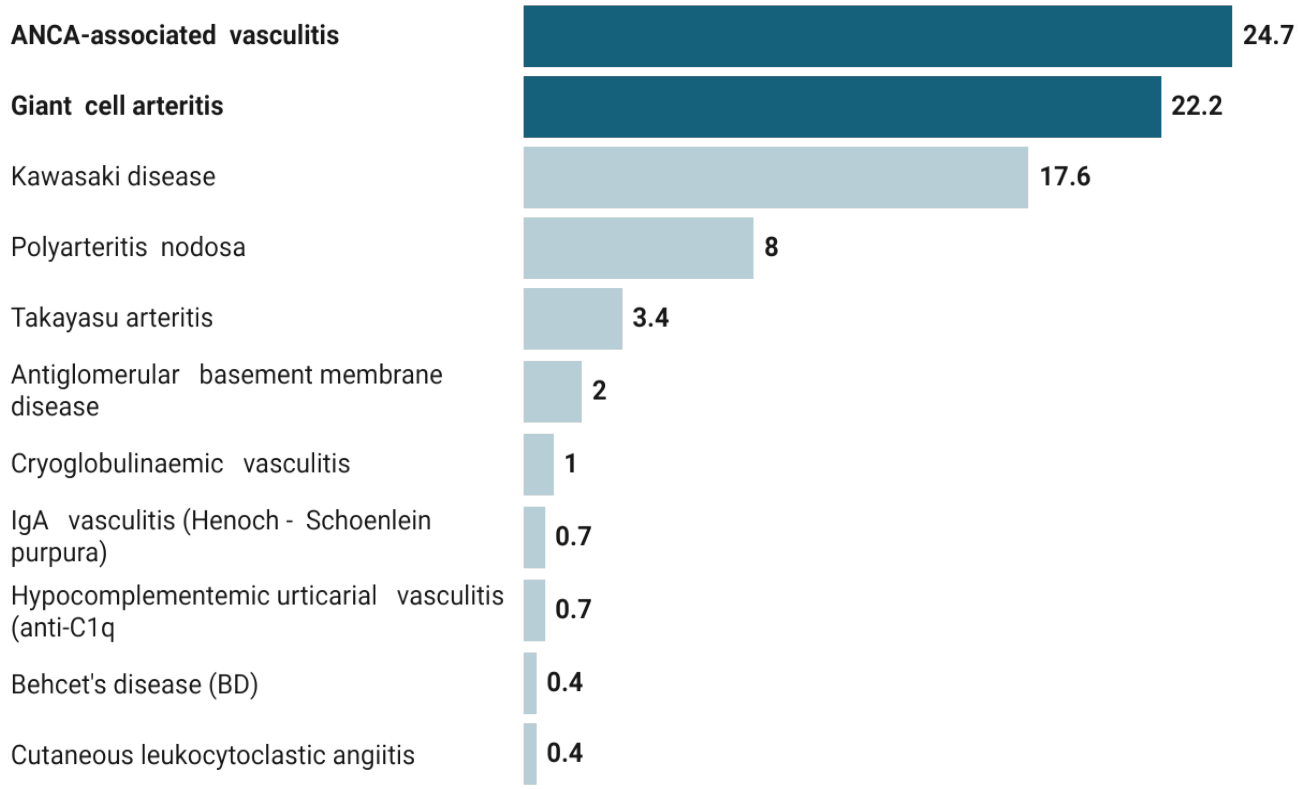



**Figure 1.2:** Maximum recorded incidence of vasculitis in routine health datasets in Europe. The numbers represented incidence per 100,000 population . The estimates were derived Watts et al.,2022 reviews on the global epidemiology of vasculitis.

### 1.3.2. The use of environmental-wide association (EWAS) studies as tools for investigating the aetiology of vasculitis

Environmental-wide association studies (EWAS) are untargeted, agnostic, and hypothesis-generating approach that explore environmental causes of poor health and diseases (79). It is an emerging approach coined by Patel et al.,2010 on the premise that putative environmental exposures identified through EWAS assessment can be used to predict and prevent environmental risk associated with disease while providing insight about disease mechanisms and possible risk factors (80). EWAs are analogous and complementary to genome-wide association (GWAS) studies, in that the risk associated with any adverse event is not only explained by polygenic genetic risk, but by external environmental factors that an individual experiences throughout their life.

Unlike GWAS, where single nucleotide polymorphisms (SNPs) are homogenous in its categorisation (i.e.: SNPs being embedded in 23 chromosome pairs), EWAS studies are complex because environmental factors have significant variation across geographies and time including an inability to adjust for a range of myriad of potential confounders that cannot be measured, for example, human mobility, and sources of exposure and personalised levels of exposure across whole populations and geographies (81).

Furthermore, EWAS require large sample size registries and biobanks designed to sufficiently detect statistical significance and clinically meaningful changes in outcomes that can be attributed to one or more environmental exposures even after adjusting for confounders and multiple comparisons (82,83) - Figure 1.3. This approach has been applied in several studies evaluating the association between environmental exposures (lifestyle, air quality and neighbourhood-level measures of social and economic status) and cardiometabolic syndromes (80,84–86). Applying this approach to rare disease research has the potential to bring about new evidence showing the role of environmental aetiologies in the onset of vasculitis, as well as allow for formation of new hypothesis of causation.

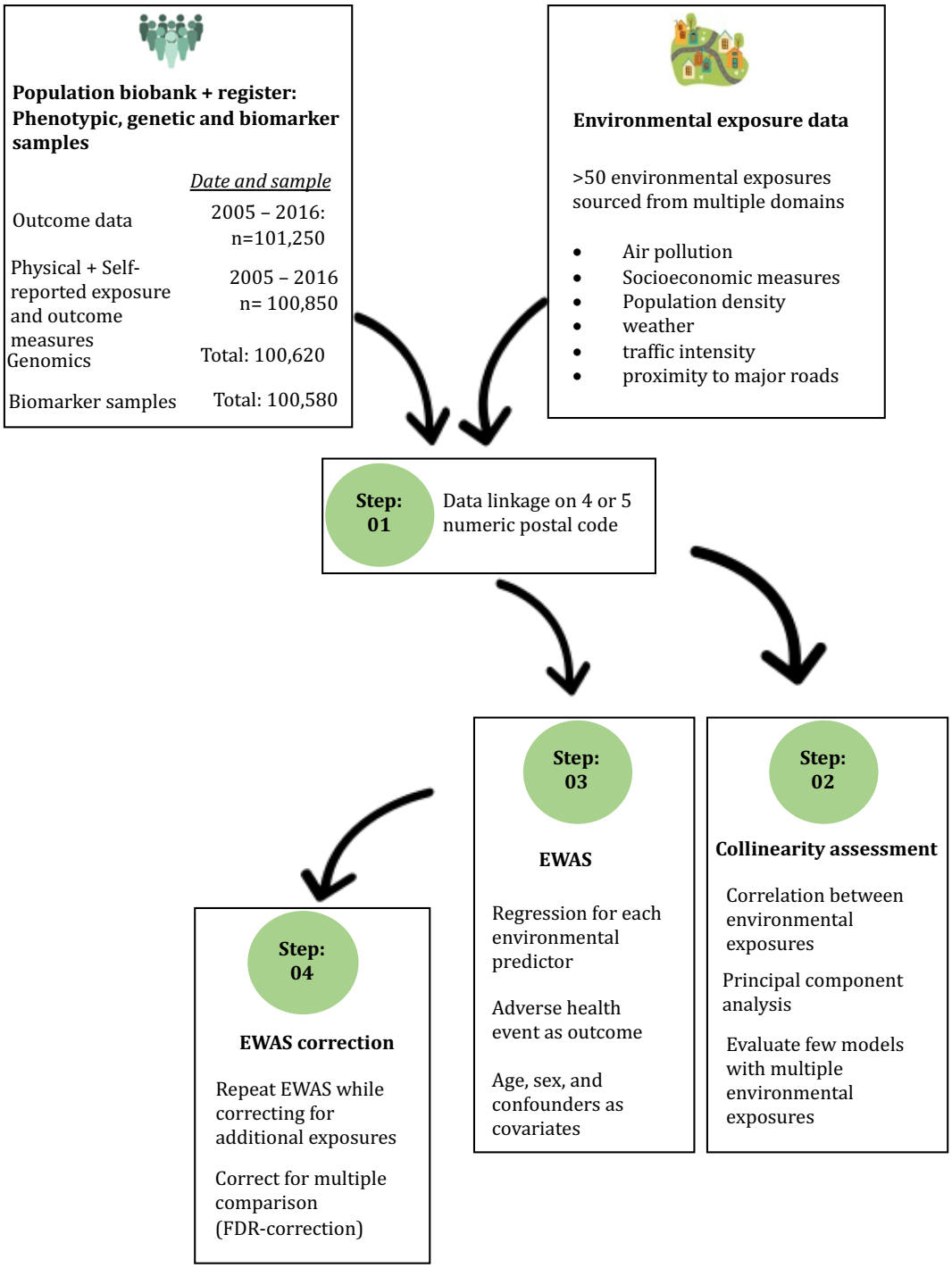



**Figure 1.3:** A schematic example of one of the approaches that can be used to conduct an environmental-wide association study (EWAS).

### 1.3.3. Section summary

This section made a case for the use of AAV and GCA as the two important vasculitis that can be used to investigate potential environmental risk factors of vasculitis. Their growing presence in routine health datasets across Europe means more research can be done to try to understand the impact of environmental triggers of vasculitis. Importantly, a case is made for the use of environmental-wide association method as a tool for investigating potential risk factors associated with vasculitis. This method is under-utilised in rare disease research, in part because of the lack of awareness, limited access to environmental variables that are meaningful as well as access to large enough sample size that can allow for power to detect significant effect of environmental exposure and onset of rare diseases. The next section will describe epidemiology of AAV and GCA globally and across the UK. It will provide empirical evidence of currently known and potential risk factors associated with vasculitis and will discuss this in terms strength, weakness and gaps from the existing literature and how this thesis aims to address them.

## 1.4. ANCA-Associated vasculitis

### 1.4.1. Definition and clinical manifestation

ANCA-associated vasculitis is defined as group of overlapping disorders that are characterised by necrotising inflammation of small vessels with few or no immune deposits(68). The vessels most affected are capillaries, venules, arterioles, and small arteries. AAV is divide into 3 clinical disease phenotypes: granulomatous with polyangiitis (GPA), microscopic polyangiitis (MPA) and eosinophil granulomatous with polyangiitis (EGPA)(70).

Each of these AAVs share non-specific clinical features of systemic inflammation, such as weight loss, malaise, fatigue, arthralgia, and myalgia including circulating ANCA targeted at protein antigens, proteinase 3 (PR3) and myeloperoxidase (MPO). More specifically, GPA and MPA tend to have overlapping features that can affect small vessels in any organ but mostly affect the upper and lower respiratory track and kidneys (Table 1.3).

Many have suggested that they should be classified based on their ANCA specificity (PR3 ANCA disease vs MPO ANCA disease). The CHCC 2012 advocates adding prefixes to AAV diseases based on the ANCA subtype (e.g.: PR3/MPO ANCA+ GPA) (22). Regardless of these recommendations, GPA is predominantly associated with PR3 ANCA while MPA is usually associated with MPO-ANCA. MPA affect small vessels with some presentation of necrotising arteritis involving small and medium arteries.

Clinically, MPA presents as a renal disease with some manifestation that are analogous of GPA and EGPA but without granulomatous inflammation (87). Other common features include Inflammation of the glomeruli and pulmonary capillaries. GPA often presents as sino-nasal disease with lower respiratory tract involvement that includes pulmonary haemorrhage and granulomatous inflammation, and glomerulonephritis (88).

**Table 1.3:** Clinical features of ANCA-Associated vasculitis and the rate of ANCA specificity

| | **GPA** | **MPA** | **EGPA** |
|---|---|---|---|
| Organs involved | Nose and sinuses, upper and lower airways, kidneys, joints, eyes | Upper and lower airways, kidneys | Upper airways, lungs, peripheral nerves, heart, skin |
| Common features | Fever, weight loss, malaise, and fatigue; ear nose and throat (ENT) signs and symptoms - nasal and oral ulcers and crusting, hoarseness, cough, dyspnoea, haemorrhage with granulomatous inflammation on biopsy. Urinary abnormalities, elevated serum creatinine, rapidly progressing pauci-immune glomerulonephritis on biopsy, purpura, and ulcers) | Fever, weight loss, malaise and fatigue, ENT signs and symptoms like GPA but without granulomatous inflammation on biopsy and less frequent. rapidly progressing necrotising pauci-immune glomerulonephritis and necrotising leukocytoclastic vasculitis. | Fever, weight loss, malaise, fatigue, and lymphadenopathy. Asthma, ENT involvement also very common, cardiomyopathy involvement in EGPA contributes greatly to mortality, serous otitis media, allergic rhinitis, nasal obstruction, recurrent sinusitis, and nasal polyposis |
| Frequency of ANCA positivity | | | |
| MPO ANCA+ | 20% | 60% | 30-40% |
| PR3 ANCA+ | 75% | 25% | <10% |

Lastly, EGPA is an eosinophilic-rich and necrotising granulomatous inflammation that affect small to medium vessels and often involving respiratory tract. About 25% of EGPA patients who are ANCA positive have no renal disease and all those who present with necrotising glomerulonephritis are ANCA positive (89).

### 1.4.2. Diagnosis

There are no diagnostic criteria for AAV. Classification criteria provided by the ACR 1990 (16,18–20), CHCC 2012 (22), and EMA 2007 (27) are used to guide assessment of disease pathology and inform early and better treatment strategy in affected patients . As initial manifestations are diverse and often non-specific, AAV can remain undiagnosed for months or years until ANCA testing is performed. Clinically, ANCA testing serves as a good classification tool for early screening and detection of disease, together with the use of X-ray, CT scan, ultrasonography guided percutaneous biopsy, haematuria and or proteinuria testing (22). Chest X-ray and CT scan helps dissect underlying pathology in patients with pulmonary symptoms. Both tools vary in their sensitivity (90). For example, CT scans are sensitive in detecting alveolar opacities as well as masses in the retro-orbital space, paranasal sinuses, and the mastoids, while X-ray are sensitive in giving broad picture about lung involvement (90). Tissue biopsy of the kidney, lung and nasal cavity are also important in establishing diagnosis. A high diagnostic specificity of ANCA ELISA (MPO or PR3) for a positive AAV may reduce the need for biopsies. Surrogate markers of disease activity (C-reactive and erythrocyte sedimentation rate) also contribute to overall diagnosis although limited in specificity and sensitivity (27,91).

In practice, when AAV is suspected, ANCA testing is done first using indirect immunofluorescence (IFF) and enzyme-linked immunosorbent assays (ELISA) for PR3 and MPO ANCA (92). Approaches to ANCA testing are provided in the consensus statements, although this varies by regions and clinical practice. Diagnosis should be reviewed periodically, particularly in cases where not all disease manifestations are consistent with AAV or when there is poor response

to treatment. A key challenge to diagnosing AAV is in its heterogeneity of symptom, and the need for dedicated clinical specialists, including laboratory medicine. Patient with AAV should therefore be assessed and managed in multidisciplinary team and early referral to specialist centre experienced in diagnosing vasculitis can improve early diagnosis and assessment. Initial assessment of AAV often targets disease activity, damage, prognosis, function, or quality of life (4).

Validated tools such as the Birmingham Vasculitis Score (BVAS), the Disease Extend Index (DEI) and the Vasculitis Damage Index (VDI) are used to evaluate disease activity and severity (93). BVAS is used to score disease activity based on 66 clinical features that are divided into 9 organ systems. Each item has a numeric score organised according to clinical relevance. Items that are scored are only attributable to disease activity (5,6). BVAS is based on clinical judgement. Difficulties arises when there is a need to distinguish between ongoing active vasculitis and symptoms due to scars without active disease. VDI assesses the extent of disease damage by using a cumulative score describing long-term outcomes for vasculitis. It includes 64 items covering 11 organ systems and defines damage as an irreversible scar present for longer than 3 months(7). The damage recorded needs to occur after the vasculitis diagnosis, but do not need to be attributable to diagnosis (e.g.: might be related to treatment). DEI is validated against BVAS and scores the number of organs affected by the small to medium vessel vasculitis. It is calculated as a subset of BVAS items and complements the overall score(8,9). Together these tools inform better management and care of patients while giving insight about extent of disease pathology.

### 1.4.3. Epidemiology

#### 1.4.3.1. Global

The incidence of AAV has increased over the past four decades improved recognition especially in regions like Latin America that historically has limited

data on disease prevalence and dynamic. To date, most of the studies reporting on the epidemiology of AAV come from Europe and North America (51).There is a growing number of studies from Southeast Asia (51,94), South America and middle east but none from Sub-Sahara Africa. Early studies, mainly from Sweden indicated that GPA may be more common in Northern Europe than MPA while countries in Southern Europe observe fewer cases (52). Figure 1.3 provides a summary of the global incidence of AAV from studies published from North America (49,71), Northern Europe (25,52,72–77), Southern Europe (53,78,79), Middle East (80,81), Southeast Asia (50,82), Oceania (83,84), and South America (85).

Furthermore, data from a recent systematic review on AAV incidence shows that GPA is indeed more common in countries from northern hemisphere with the highest incidence being recorded Minnesota, US and while lower incidence is seen Italy Turkey (71) – Figure 1.4. Overall, the global incidence of AAV is 17.2 per million person-years and a prevalence of 198 cases per million persons with variation in incidence that ranges from 8.1 per million person years to 33.0 per million person-years and for prevalence, 44.8 case per million persons to 421 .0 cases per million persons. Incidence is slightly higher in men than women and in those over the age of 50 years old being. When assessing the incidence by AAV subtype, GPA accounts for the highest incidence (9.0 per million), followed by MPA (5.9 per million) and EGPA (1.7 per million). When assessed by geography, GPA and MPA have a higher occurrence in North America while MPA occurs more in Asia in countries such Japan and China (71). In all countries reporting on AAV, EGPA has the lowest incidence with the ranges that starts from 0.14 to 4.0 cases per million persons per year and no evidence of temporal variation. In general, GPA and MPA is more common in white population, with recent data indicating potentially higher frequency in minority Chinese Dong population. Further variation in incidence by ethnicity have been examined in fewer studies, with one major study being from the UK and is further discussed below.

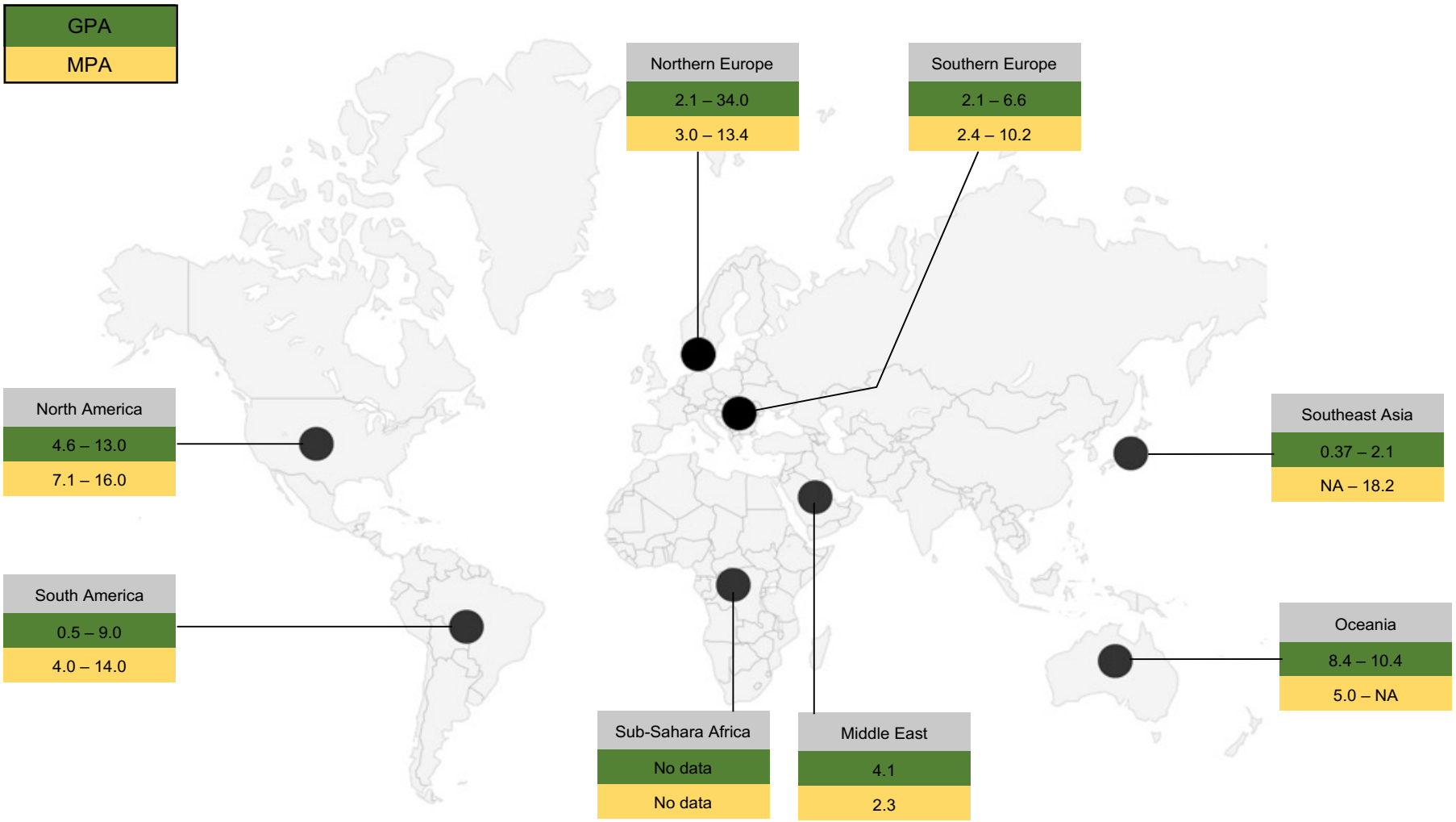



**Figure 1.4:** Global incidence of ANCA-Associated vasculitis by continent and summarised as incidence per million persons 50 years and older.

### 1.4.3.2. United Kingdom

The epidemiology of AAV in United Kingdom has also shown a steady increase over the years, with estimated prevalence being expected to increase by 35% in 2034 for areas such the Midlands of UK (95). Historic estimates show the prevalence of AAV increased by more than two-fold between 1990 and 2005 (28.8 per million to 64.8 per million population) (95). This increase is again owed to improved classification criteria and awareness by physicians in the UK. Like global pattern of disease incidence, AAV appears to be more common in men than women, with mean age of onset being 55 years and older. Data reported from the UK General practice research database revealed the overall incidence of AAV was 8.4 cases per million population for GPA, 5.9 per million population for MPA and 0.9 cases per million population for EGPA (96). Estimates reported from the Scottish renal biopsy registry, Norwich Health authority, medical records from Nottingham derby urban area and Cheshire County shows some temporal variability in incidence and prevalence of AAV, and higher incidence being seen the Midlands of the UK (Table 1.5)

**Table 1.4:** Incidence of ANCA-Associated vasculitis in UK

| | Incidence per million population over 50 years old | | | | | |
|---|---|---|---|---|---|---|
| Location | AAV | GPA | MPA | EGPA | Latitude | Ref |
| Norwich | 18.3 | 10.6 | 8.4 | 3.1 | 52.6278 | (96) |
| Nottingham | 23.1 | 8.2 | 13.4 | 1.5 | 52.9540 | (95) |
| West Cheshire | 1.2 | NR* | NR | NR | 53.2303 | (97) |
| Scotland | 15.1 | 5.8 | 9.3 | NR | 56.4907 | (98) |

*NR – not reported.

Overall, the incidence rate in Nottingham was AAV 18.3 million case per million population (1990 - 2005), 23.1 per million population (2007 to 2013) in Nottingham. In West Cheshire, the incidence was 1.2 cases per million West and for the Scottish renal biopsy (2014 - 2018), it was 15.1 million per million population in (95,97,98).

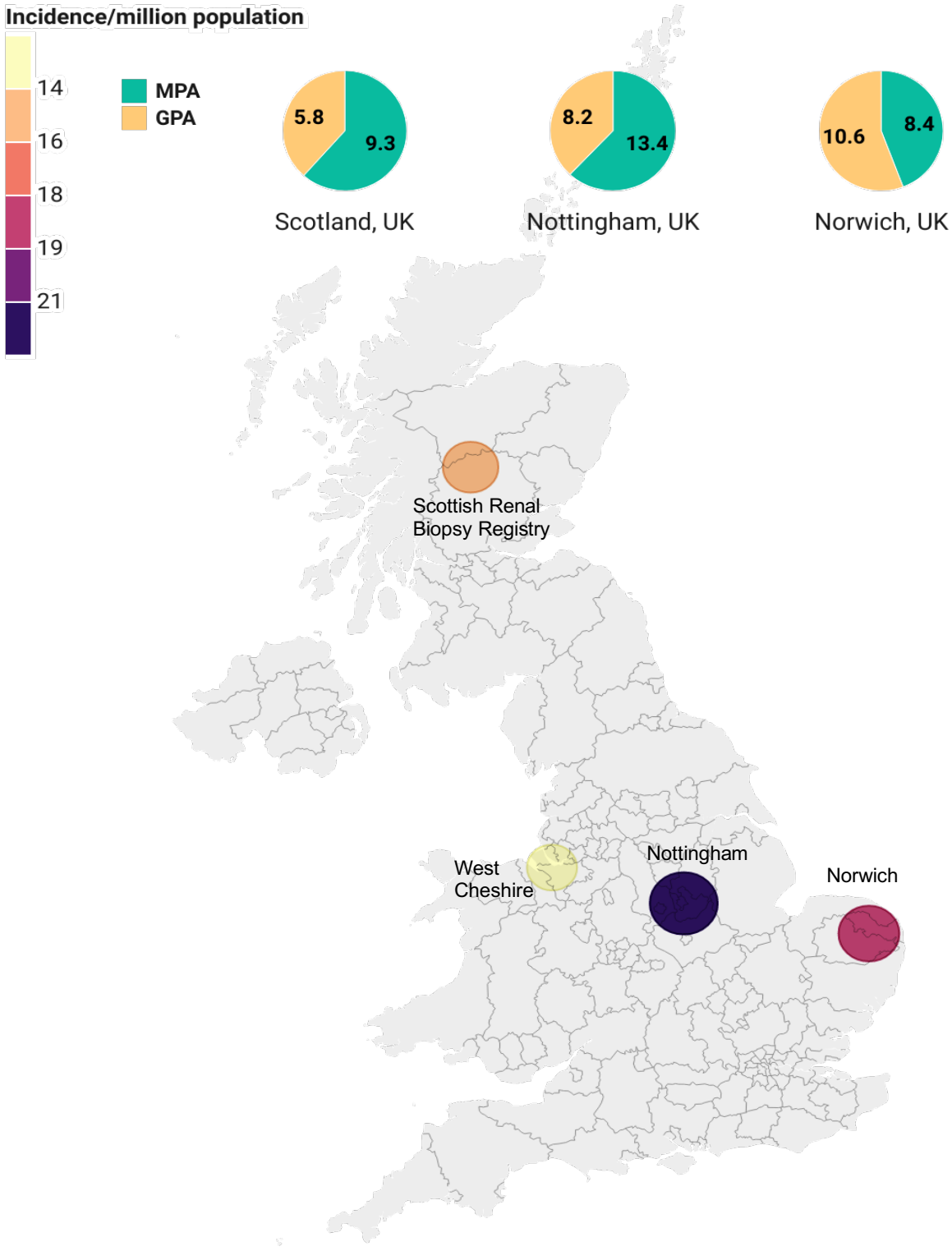



**Figure 1.5:** Incidence of ANCA-Associated vasculitis in the UK per million persons aged 50 years and older. The figures summarise incidence estimates by counties where incidence of AAV has been reported in the England and Scotland. [*Figure created with datawrapper.de]*

The incidence of AAV was lower in West Cheshire because the data were derived from a single hospital centre. For the rest, the incidence was estimated using administrative health data, with census population being used as a population denomitor. Additionally, MPA appear to be more common in the midland of UK, while GPA appears to be higher in the north and east of UK (Figure 1.4). This variation in incidence can be explained by differences in methodology including period of study, data source and geographic coverage. The role of environment in AAV onset and incidence is not clear and has not been studied in the UK. Addressing this gap will provide insight about the social, demographic, and environmental determinant of vasculitis .

### 1.4.4. Pathogenesis

The pathogenesis of AAV is complex and includes genetic, environmental exposures. Genome wide association studies, and in vitro and vivo animal studies have identified several genetic associations that suggests that major histocompatibility complex (MHC) and non-MHC genetic loci are distinct to PR3 and MPO may increases susceptibility of AAV (99). Specifically, polymorphisms in genes coding HLA-DP, alpha-1 antitrypsin (SERPINA1 and PRTN3) and HLA-DQ are linked with the development of PR3-AAV and MPO-AAV (100). Environmental factors that will be discussed later (e.g.: Crystalline silica, smoking, infection) are potential initiators of pathogenic pathways that lead to the chronic and granulomatous inflammation of AAV in susceptible individuals (101–104). The exact mechanisms by which genes and environment interact to drive the onset of disease are not clear. Current evidence suggests molecular mimicry and the priming of dendritic cells might play a role in inducing a cascade of events that lead to increased activation of neutrophils and monocytes, as well as the production of auto reactive CD4 and CD8 T cells that promotes the production of ANCAs, that in turn activate the alternative complement system (ACP), associated with kidney injury in GPA and MPA (105). One way which ANCA-activated neutrophils drive disease pathology is through C5a alternative pathway that involves priming of neutrophils and inflammatory interaction with endothelial cell layers of affected vessels (68).

Drugs targeting that inhibit the activation of C5a receptor (for e.g.: avacopan) and B cell response have shown to be efficacious in managing patients with AAV (106). Additionally, CD68+ CD163+ macrophages also play an important role in initiating glomerular injury and fibrosis and are frequently found at the site injury (69).

## 1.5. Giant cell arteritis

### 1.5.1. Definition and clinical manifestation

Giant cell arteritis is defined as a granulomatous inflammation of the blood vessel wall and a maladaptive immune response to injury that promotes intimal hyperplasia, adventitial thickening, and intramural vascularisation, that ultimately threaten vessel integrity and tissue perfusion (107). GCA affects medium and large vessels, usually the aorta and/or its major branches, particularly branches of the carotid and vertebral arteries and with temporary artery involvement, although not in all cases (67). The term temporal arteritis, previously used to describe GCA is therefore not a suitable alternative naming as other forms of vasculitis can also affect the temporal artery. In most cases, temporal arteritis is still a common feature of GCA and may present with broad spectrum signs and symptoms attributable to systemic inflammation and ischaemia (108). The main clinical features of GCA are categorised into systemic or tissue/organ specific symptoms related to ischaemia. Systemic symptoms are broad and non-specific and include weight loss, anorexia, arthralgia, fatigue, lethargy, low grade fever, and myalgia. Organ and tissue specific symptoms related to ischaemia include headache, neck pain, jaw and tongue claudication, neck pain and neurological deficit, vision disturbance, tender scalps, neurological deficit (Table 1.4) (67,109). Nomenclature of GCA has evolved to reflect changes in the naming and the involvement of broad large and medium vessels and inflammation, with terms like large-vessel GCA (LV-GCA), cranial GCA (C-GCA) and LV with cranial involvement now being suggested (67). Patients with LV-GCA are generally younger at presentation, more likely to be female and have bilateral arterial involvement than those with

C-GCA (110). 20-80% of with symptoms involving large vessels, have clinical manifestation of upper and/or lower extremity vasculitis that precede diagnosis by up to 12 months (67). In these cases, there is visible unilateral or bilateral manifestations of upper and or lower extremity claudication (111). Aortic involvement is often clinically silent or may manifest as systemic inflammatory syndrome, until aneurysm is discovered during routine chest radiography or serious complications (112).

**Table 1.4:** Clinical manifestation of large-vessel GCA (LV-GCA) and Cranial GCA (C-GCA)

| | Large-vessel GCA | Cranial GCA |
|---|---|---|
| Headache | ± | ++ |
| Fever, weight loss | ±+ | + |
| Temporal artery swelling/ tenderness | + | + |
| Extra-cranial artery bruit | ++ | - |
| Visual symptoms or complications | - | ++ |
| Limb claudication/and or blood pressure discrepancy | ++ | - |
| Polymyalgia symptoms | ++ | + |
| Acute phase reactants | + | ++ |
| Peripheral arthritis/RS3PE syndrome | ± | ± |

### 1.5.2. Diagnosis

In general, there are no official diagnostic criteria for GCA, only classification criteria (ACR, CHCC) designed to distinguish patients with GCA from patients with other vasculitis (22,113). The American College of Rheumatology traditional criteria for giant cell arteritis recommends that patient may be classified as having GCA if at least three of these five criteria are present:

- Age of onset >50 years
- New onset of localised pain in the head

- Temporal artery tenderness to palpation or decreased pulsation, unrelated to arteriosclerosis of cervical arteries.
- Elevated erythrocyte sedimentation rate (ESR)
- Biopsy specimen with artery showing vasculitis characterised by a predominance of mononuclear cell infiltration or granulomatous inflammation, usually with multinucleated giant cells.

Temporal arterial biopsy is particularly important and serves as the most definite evidence of GCA by identifying the typical histological features such as inflammatory infiltrates in adventitia and media of arterial wall that may or may not have the presence of giant cells in the elastic lamina fragments) (109). Imaging techniques like angiography are used as reference standards, however new radiation-based imaging is increasingly being used (for e.g.: computed tomography (CT), CT angiography, magnetic resonance imaging (MRI) and magnetic resonance angiography (MRA(114). MRI can be complemented by MRA to detect extra-cranial involvement, however more data is needed to support this method (115). Ultrasonography can also be used to examine large-vessel involvement, mostly crania but does not allow evaluation of thoracic arteries (116). 18F-fluoredeoxyglucose positron emission tomography (FDG-PET) is a promising diagnostic tool for extra-cranial GCA and can detect early signs of disease in its preclinical phase (117). However, it's limited in that there is no standardised method to quantify the dose FDG uptake and may not be suitable for temporal or cranial arteries.

### 1.5.3. Epidemiology

#### 1.5.3.1. Global

GCA is the most common primary vasculitis with estimated global prevalence of 51.7 cases per 100,000 people over the age of 50. This number is expected to grow by 2050 owing to a growing ageing population in developed countries (118). Specifically, it is estimated that more 3 million people will be diagnosed with GCA in Europe, North America, and Oceania. The extent to which similar pattern will be seen in other regions and continent is not yet clear. Currently,

there is substantial variation in the incidence of GCA ranging from 0.34 cases per 100,000 persons (Hong Kong) to 76 cases per 100,000 persons (Denmark) (44). The exact estimate varies depending on case definition criteria (biopsy proven cases vs diagnostic coding or classification criteria). To date, the global pooled incidence of GCA is 10 per 100,000 persons over the age of 50 years old. The global pooled incidence of GCA is 10 per 100,000 persons over the age of 50 years old(44). The incidence appears to vary by latitude with Scandinavian countries accounting for the highest incidence and Southeast Asian countries the least (Figure 1.6). When assessed by continent, there is considerable heterogeneity between and within continents. This is particularly true for countries in northern Europe. Regional incidence ranges from 2.3 to 76.6 cases per 100,000 persons in Northern Europe, 2.2 to 14.1 cases per 100,000 for Southern Europe, 0.49 to 11.3 cases per 100, 000 person for Middle east and North Africa, Sub-Sahara Africa, 0.34 for Southeast Asia, 8.6 for South America, 1.02 to 19.8 in North America, and 3.2 to 12.73 for Oceania. Overall, historic assessment of GCA incidence show that there has been a decrease in the number of new cases with regions with a growing aging population showing modest increases in incidence (119,120).

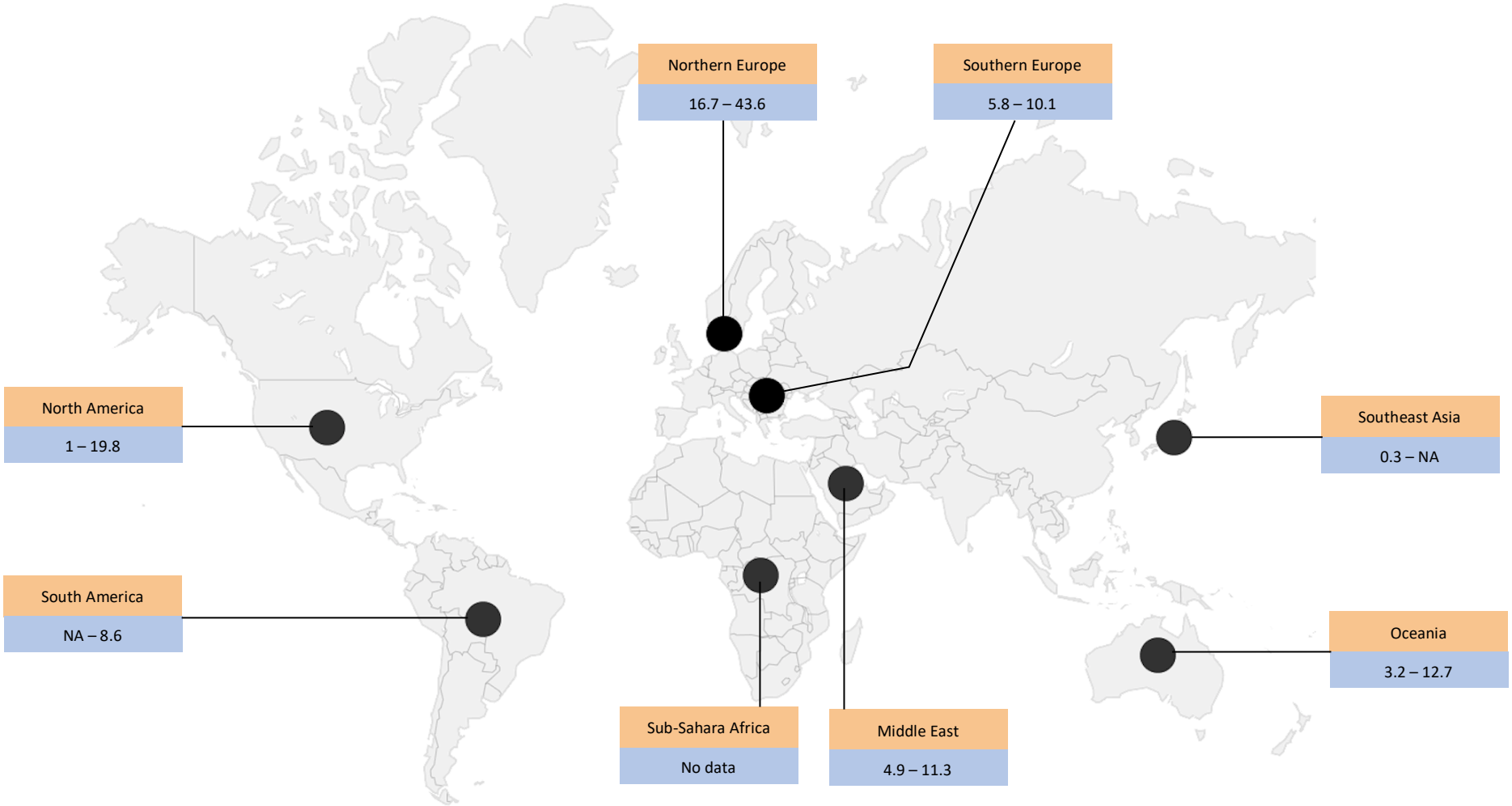



**Figure 1.6:** Global incidence of Giant cell arteritis by continent and summarised as incidence per million persons 50 years and older. The estimates are derived from a recent systematic by Li et al.,2021

### 1.5.3.2. United Kingdom

The incidence of giant cell arteritis in the UK appears to be lower compared with most Northern European countries, particularly Scandinavian countries, and Germany (44,121–125). This variation in incidence could in part be explained by the fact that majority of epidemiological studies in Europe come from hospital records with much bigger surveillance and geographic coverage, unlike the UK.

In the UK, most studies on the epidemiology of giant cell arteritis come from primary care records (Table 1.5). There is limited data from tertiary care/ hospital admission records where most of the patients are likely to be diagnosed. To date, current estimates show that the incidence of GCA ranges from 0.4 to 4.31 case per persons over the age 50 years old (121–125).

**Table 1.5:** Incidence of giant cell arteritis in the UK population 50 years and older

| Source | GCA cases (n) | Setting | GCA per 100,000 persons | Time period | Reference |
|---|---|---|---|---|---|
| UK Clinical Practice Research Datalink (CPRD) | 4,671 | Primary care | 1.0 | 2000 - 2011 | Petri et al., 2015 (122) |
| Norfolk primary care record | 21 | Primary care | 0.4 | 2013 | Yates et al., 2016 (125) |
| NHS England Hospital Episode Statistics | 7,864 | Tertiary care | 4.31 | 2002 - 2013 | Mollan et al., 2015 (123) |
| UK General Practice Research Database (GPRD) | 3,928 | Primary care | 2.2 | 1990 - 2001 | Smeeth et al.,2006 (121) |

Incidence rate is particularly higher in women aged 70 to 79 years old in the UK, with some data to indicate that the overall trend is growing over time,

initially at the primary care level (1990 - 2001) and later tertiary care (2002 to 2013) (121,125). This growth can again be attributed to the introduction of classification systems and potentially, an ageing population. It is not clear whether there is variation in incidence of GCA that is attributable to geography and overall environment. Smeeth et al., 2006 attempted to show this by highlighting variation in GCA incidence across all UK geographies. From this assessment, there is some indication that there might be higher incidence rate in the Eastern and Southeast region of the UK (Figure 1.7) (121). Whether this variation is true at tertiary care level is not yet clear. Understanding the role of the environment deciphering the epidemiology and biology of GCA biology is essential as the prevalence continues the to grow, potentially allowing for imp


**Figure 1.7:** Age standardised incidence ratio (with 95% confidence interval) of giant cell arteritis. A map extract from Smeeth et al.,2006.

### 1.5.4. Pathogenesis

The past two decades have seen a growing number of studies implicating several mechanisms associated with the pathogenesis of GCA and that contribute to the loss of immune tolerance and disease pathogenesis that are preceded by ageing, genetic and environmental factors (126). The exact aetiology associated with the onset of GCA is not yet clear.

Evidence from genome-wide association studies have identified several links between genetic polymorphism of HLA-DRB1*04 alleles, specifically DRB1*0401 and DRB1*0404, and an increased risk of GCA(127). Age and immunosenescence process that comes with ageing which is characterised by a reduction of naive T cells and T reg, and increase in inflammatory cytokines (TNF, IL-6, IL1-B) including a slow cellular immunity to inflammatory signals may play an important role (67,128,129).

Seasonal infections and geographic exposures in combination with ageing, may drive epigenetic changes in susceptive individuals, thereby increasing their risk of GCA onset(130,131). In general, it is this interplay between ageing, epigenetic genes and environment that contribute to the loss of immune tolerance of the arterial wall and an innate immunological pathway involving both the innate and adaptive immune response (67) .

This process involves several mechanisms. One: is the loss of anti-inflammatory regulatory T cells involved in the regulation and suppression of pro-inflammatory T cells in lymph nodes. Secondly, the dysfunctional regulation of PD1 or PDL1 inhibitory pathways contribute to the inflammatory response of the artery and activates pro-inflammatory T cells and macrophages and accelerated hyperplasia (132). Thirdly, damage of the arterial is driven by the leakiness of the endothelial barrier, which prevents migration of circulating cells in the vessel wall. Specifically, circulating monocytes produce excess metalloproteases (MMPs), digests the subendothelial basal lamina layers and enable T cells, which are also independently capable of MMP2 and MMP9 production and infiltrates (133). Lastly, patients with GCA present with high

concentration of circulating immature neutrophils in peripheral blood enriched with reactive oxygen species. This enables them to breach the endothelial barrier and propagate further inflammation through leukocyte-endothelial cell interaction.

## 1.6. Genetic and environmental risk factors associated with ANCA-Associated vasculitis and Giant cell arteritis.

### 1.6.1. Genetic risk factors

Evidence of important genetic association of AAV and GCA have been growing over the past decade and includes evidence of familial association. The most convincing evidence has been with major histocompatibility complex (MHC) and non-MHC genes. Polymorphisms in genes coding for HLA-DP, alpha-1 antitrypsin have particularly shown to be associated with the development of AAV (100,134–136).

Variants in genes that encodes for the protein tyrosine phosphatase non-receptor type 22 (PTNP22) have also shown significant association with AAV (100). PTPN22 affects the responsiveness of B and T cell receptors and has been implicated in multiple autoimmune diseases including systemic lupus erythematosus, rheumatoid arthritis, suggesting that genetic risk factors common to other autoimmune disease also apply to AAV (137). Furthermore, there is a genetic distinction between GPA and MPA that are associated with ANCA specificity and suggest that immune mediated response against PR3 may be central pathogenic feature of PR 3 AAV (100).

Similarly, polymorphism in genes that lie within HLA-class II region are associated with the development of GCA. Specifically, HLA-DRB1*04 alleles gene which is most evident in people both from Scandinavian countries where GCA is highly prevalent and from the Southern region of Europe (Spain and Italy) (138). Besides the HLA-class II encoded genes, it is likely that other genes contribute to the susceptibility of these GCA and other overlapping inflammatory

autoimmune diseases. For example, polymorphism of genes coding cytokines including their receptors (TNF-alpha, IL-6, IL-1, I-CAM-1) implicated in the pathogenesis of GCA plays an important role in increasing the susceptibility of disease onset (67,127,139).

## 1.6.2. Environmental risk factors and exposures

### 1.6.2.1. Definition

There is no single definition to describe environmental exposures. The current definition provided by the national institute of health (NIH) states that an environmental exposure is any "*chemical, biological, or physical substances found in the air, water, food, or soil that may have harmful effect on a person's health*" (140). These substance can be measured directly or indirectly (141).

Environmental protection agency guideline on exposure assessment (US EPA,1992) highlights that direct exposure assessment is an "*evaluation of an exposure as it occurs, by using direct methods to measure the substance concentration at the interface between the person and the environment as a function of time"* (142). Direct or 'point of contact' exposure assessment is achieved through personal monitoring campaigns conducted using personalised air samplers (e.g.: for temperature or air pollution) or laboratory measurements (e.g.: for viral or bacterial infection or drug dose) (143). These measurements can be continuous or fixed and aim to capture the exposure dose concentration at a given time. It is important to note that direct exposure assessment can vary in accuracy and may require extrapolation of short-term sampling for long-term exposure.

In contrast, indirect exposure assessment "*quantifies exposures by estimation of both the amount of substance contacted, and the frequency/duration of contact, and subsequently link these together to estimate exposure or dose*" (144). This approach uses indirect or broad estimation of dose and involves quantitative values as an input through the development of exposure scenarios.

Example of this approach include estimation of occupation pollutant through exposure matrices. Other examples include estimation of food, alcohol, and tobacco intake based on self-reported questionnaires or geographic estimates related to distance measures from areas with high air pollution.

For vasculitis, most studies reporting on the aetiology of vasculitis uses indirect exposure assessment tools and measures. For example, these measures cover a myriad of environmental factors that include serious infections that precede the onset of disease, the number of cigarettes consumed over a given period ( a measure of smoking), reaction to medication, occupational exposures, and environmental dusts, including geographic exposures associated with rurality and seasonality as proxy measure of exposures associated with seasonal change (Figure 1.8).

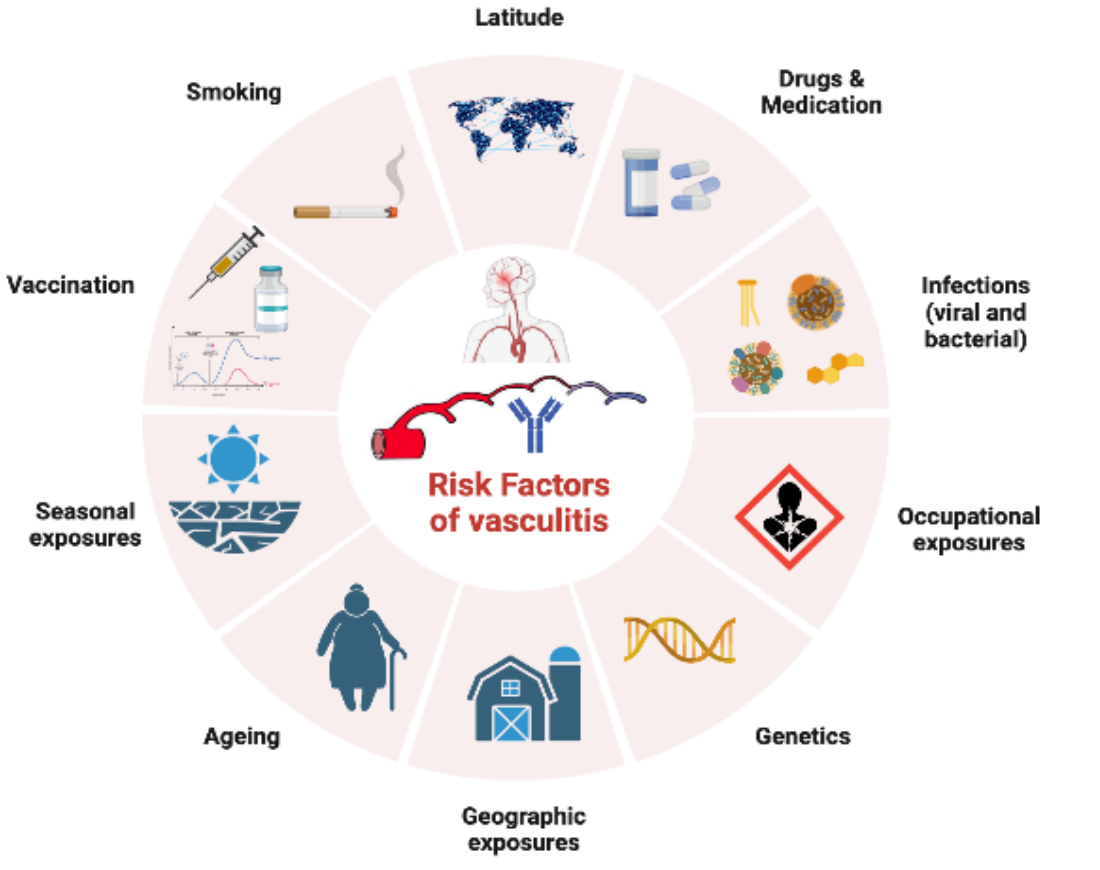



**Figure 1.8:** A schematic overview of current risk factors associated with vasculitis.[*Figure created with biorender.com.]*

Studies reporting on the relationship between these exposure and onset of vasculitis are discussed in following paragraphs. Overall, it is not clear which of these factors have causal association with vasculitis. There is large heterogeneity in the methods used for indirect exposure assessment, in terms of accuracy and precision of exposure and size. This section provides a summary of the method, the strength of results including limitation of the results.

### 1.6.2.2. Geo-epidemiology, Ultraviolet B, and vitamin D

There is strong evidence to show that the occurrences of autoimmune diseases, like rheumatoid arthritis and systemic lupus erythematosus, follow a north-south gradient, with countries in the northern hemisphere reporting higher prevalence compared with southeast and south-west countries. In the case vasculitis, studies on geo-epidemiological factors associate with the onset of disease suggests that latitude and possibly a lack of ultraviolet B (UVB) in preceding vitamin D may be associated with an increased risk of ANCA vasculitis and GCA (145–147).

For example, studies on the epidemiology of AAV suggest that latitude may explain variation in ANCA subtype between global north and south including within country or regional incidence of AAV. Specifically, GPA appears to be more common in people of European ancestry than those of Asian background (148). Evidence from the DCVAS study suggests that MPO-ANCA is more common in Japanese, Chinese and Southern Europeans compared with northern Europe and America (149). Additionally, a recent study from China reported the frequency of AAV increased with latitude. Scott et al., 2022 showed that residential latitude was positively correlated with AAV relapse potentially due to a lack of ultraviolet B exposure in winter (147,150)

In the case of GCA, a recent meta-analysis has shown that latitude is significantly correlated with incidence of GCA but not prevalence or mortality (44). This correlation may be explained by the lack of UVB exposure between the northern and southern hemisphere. For example, Wing et al., 2015 reported

that the incidence of GCA was correlated with geomagnetic activity and other surrogate measures of the solar cycle and UVB exposure(151). It is not clear whether the effects of latitude and UVB measures are a function of vitamin D status and if so, what extent this is true at a population level. Currently, more studies are needed to provide further clarity. Specifically, to characterise the role of cumulative vitamin D dose and concentration and its link with vasculitis risk.

### 1.6.2.3. Seasonality

It is not clear whether the epidemiology of vasculitis has a seasonal or periodic cycles that may give insight about its aetiology. Unlike vasculitis in children where the incidence of Kawasaki disease has shown to follow a seasonal and periodic cycle that may be associated with tropospheric wind, and potentially regional microbial agents (152), the incidence of AAV and GCA do not seem show clear pattern in this regard.

For example there is data to suggest that the incidence of GCA is associated with warmer seasons, specifically spring or summer (153–156). It is worth noting the higher frequency of GCA cases in these season do not show statistically significant variation from other seasons, in part of because of the low sample size from these studies and considerable heterogeneity between studies in terms of quality, classification criteria and definition of disease. For AAV, the evidence on seasonality is mixed with some studies reporting higher frequency of AAV in warmer and cooler months (157–160). Chapter 2 will use a systematic approach to assess this and to further understand the impact of seasonality and potentially weather conditions and their effects on vasculitis onset.

Temporal peaks in incidence of vasculitis have also been observed in studies from Sweden and Spain (99,100). These peaks could in part be explained by cycles of seasonal infections and generally regional outbreaks of seasonal flu or infection. For example, Tidman et al.,1998 showed that in Sweden there were a series of peaks in incidence of AAV that ranged from 0.5 cases per 100 000

to 2.5 cases per 100 000 population every 3 to 5 years between 1975 to 1995 (161). Draibe et al., 2018 reported much shorter peaks in AAV incidence that ranged from 10 to 12 months between 2001 to 2014 and mainly starting February(162). There is limited data to show similar fluctuation in GCA incidence. Given such gap, chapter 2 will assess in more detail the seasonal pattern in both AAV and GCA to get further insight about the temporal cycles of vasculitis incidence and potential links with environmental risk factors

### 1.6.2.4. Infection and vaccination

There is a consensus that infection plays an important role in the onset of AAV, though this is not clear for GCA. The association between serious infection and AAV is well documented. The exact microorganisms, whether bacterial or viral, are not yet known. Upper respiratory and systemic infections seem to be the most common occurrence in patients diagnosed with AAV(103,104,163–166). Pneumonia and sepsis are commonly reported in these patients when compared with population controls (104). A recent systematic review showed that Staphylococcus aureus (SA), *Mycobacterium* spp, *Coccidioides* spp, *Rickettsia rickettsii*, nasal carriage of Staphylococcus spp, Epstein-Barr virus (EBV), Cytomegalovirus (CMV) and Dengue virus are among the top reported pathogens associated with MPO- ANCA, and often precede the onset of disease (164). It is worth noting the number of patients included in most of studies are small (n=<30), further validation studies assessing the role of potential infectious pathogen are needed to provide clarity on important antigens associated with onset of vasculitis

Additionally, vaccines are suspected to be associated with the onset of vasculitis, particularly Kawasaki disease in children and AAV and GCA in adults. Most studies reporting on the impact of vaccination on AAV are case series and case reports (167). There are only a handful of observational studies and they show no causal relationship between Influenza and Hepatitis B (HBV) vaccine and the risk for the onset of GCA and GPA (167–169). While case reports and case series studies have provided some insight into potential association

between vasculitis and vaccines and their corresponding antigens, further study will be needed to validate this in retrospective cohort studies.

### 1.6.2.5. Geography

There is growing evidence to indicate the prevalence of vasculitis may vary by urban and rural geographies. For AAV specifically, several observational studies from Canada, France, Australia, and UK have reported a higher frequency AAV incidence in rural areas (98,158,170,171). It is not clear whether this difference is true based on AAV phenotype. For example, a recent study by Aiyegbusi et al., 2021 showed that the incidence of biopsy-proven GPA was associated with rurality in Scotland (98). Anderson et al., 2013 reported similar results in the northern Saskatchewan population in Canada(171). Ormerod et al., 2008 reported higher frequency of MPA in rural Australia (rural incidence per million :13.9, 95%CI 7.7–23.5 vs urban incidence per million pop: 1.6, 95%CI 0.2–7.2), while Mahr et al., 2004 showed no significant variation in AAV incidence based on geography in France (170,172) .

Differences in AAV incidence may reflect heterogeneity in study methodology in terms of sampling strategy and recruitment. They may also reflect real differences driven by environmental exposures linked with each setting. For example, Chung et al., 2020 attempted to show this in their investigation into differences in AAV incidence based on rural and urban geographies in Australia while considering the impact of environmental exposures(173). For this, the authors used a standardised questionnaire to assess whether AAV patients from urban and rural areas had variation in environmental and occupational exposure that reflected each setting. Overall, the study reported that AAV patients from rural areas had increased exposure to silica, solvents, and dust from farming and gardening (173). It is worth noting that these results were limited by small sample (less than 70 cases) in each urban and rural geography. More work will need to validate and dissect the role of geography and environmental exposures in driving the onset of AAV. Similarly, for GCA, there is only one study reporting higher incidence of GCA in urban areas (171) and no further evidence to support variation of incidence by geography.

### 1.6.2.6. Environmental dust (organic and inorganic)

Environmental dusts are tiny air particles that are released in the air during natural erosion of soil, sand, rock, and inorganic material (174). The particles can be organic and inorganic source and often propelled after a dust storm or strong wind and after large explosions or oscillation of concrete. Inorganic dusts are a by-products of manufacturing processes like cement and coal manufacturing. In some cases, they occur after a natural disaster, such as the earthquake. Organic dusts are of plant and animal source and can consist of manure and compost including soil from farmland.

The hypothesis that environmental dusts may be associated with vasculitis was first described in two major publications that detected cases of vasculitis several years after a collapse of buildings with large quantities of asbestos, silica, and other major inorganic dusts. one such study is by Webber et al., 2015. The authors looked at the impact of inorganic dusts exposure after the collapse of the World Trade Centre (WTC) in September 2001, New York (NY)(175). The study used a nested case-control design with first responders (fire fighters, medical emergency personnel, law enforcement) who were at the site after the incident. This group served as the main population of interest. Outcome data were sourced from the NY heath district and included rheumatologist confirmed diagnoses. Dust exposure was defined as time spent working at site (acute: morning of 9/11 or shortly thereafter and chronic exposure: number of months at spent on site). Overall, 59 cases with systemic autoimmune disease were identified and included few cases of vasculitis. Each case was matched by sex, race, work assignment, and the year of hire. Of the 59, 6.8% of cases had vasculitis (n=4), most incident cases were with rheumatoid arthritis (37%, n=22), spondylarthritis (22%, n=13) and systemic lupus erythematosus (12%, n=7).For each month worked at the site with rubbles and inorganic dust exposure was associated with 13% increased odds (odds ratio: 1.13 (95% CI 1.02–1.26) ) of developing a systemic autoimmune disease.

In addition, a recent study by Ying et al 2021, reported that historic exposure to inorganic dusts via military service was associated with systemic autoimmune diseases including vasculitis (176). Proxy measures of inorganic dust exposure comprised of any interaction with explosives, being involved in firefighting, interacting with military utility, and vehicle repair.

Interestingly, majority of evidence on the impact of organic dust exposure on vasculitis come from studies reporting higher occurrences of AAV after a mass earthquake. For example, Yashiro et al., 2000 were the first to report an overall increase in frequency of MPO-ANCA vasculitis three years after the 1995 Great earthquake in Kobe (Japan). Patients seen in hospitals that were within 80km radius of the earthquake (Kobe city) had increased levels of white blood cell count (11,321cells vs 8,116 cells/uL; $P<0.05$) compared with those seen in care settings beyond the 80km radius (Kyoto city) (76). Similarly, Takeuchi et al.,2017 reported similar increases after the Great East Japan earthquake in 2011. Specifically, the authors showed that there was an increase in incidence of MPA from 17.4 cases per million population per year before the earthquake to 33.1 cases per million population after the earthquake (78). In contrast, a study conducted after the Christchurch earthquake in New Zealand reported a non-significant change in the incidence of AAV three years after the earthquake(77).

While findings from all these studies are confounded by a myriad of unmeasured factors (e.g.: social, economic, and behavioural), they support the hypothesis that organic and inorganic substances propelled in the air during natural disasters and through manufacturing in various settings may be associated with an increased risk of vasculitis in susceptible individuals.

### 1.6.2.7. Smoking

Tobacco smoking has causally been linked with several autoimmune diseases, such as rheumatoid arthritis and systemic lupus erythematosus and now recently vasculitis (177,178). Evidence on smoking and the risk of vasculitis is mainly on GCA with several studies emerging on AAV. For example, a recent

metanalysis of 13 studies reporting the association between smoking history and a risk for GCA, revealed that being a current or previous smoker was associated with a 19% (OR: 1.19, 95% CI: 1.01 – 1.39) increased odds of developing GCA(179). The meta-analysis reported large heterogeneity between studies and highlighted the need for more better designed studies that account for potential dose and time dependent risk factors.

For AAV, a recent study by McDermott et al., 2020, reported a dose-dependent response between number of years of smoking and an increased risk of AAV in US. Being a current smoker was associated with nearly 3-fold (OR: 2.7, 95% CI: – 1.8 – 4.1) increased odds of developing AAV. Previous smokers also showed 60% (OR:1.6, 95% CI: 1.3-2.0) increased odds of developing AAV compared with non-smokers (101). Overall, these associations were especially strong for participants with MPO-ANCA–positive disease (former smokers: OR, 1.7; 95% CI, 1.3-2.3; current smokers: OR, 3.5; 95% CI, 2.1-6.1) but not those with PR3-ANCA–positive AAV (former smokers: OR, 1.3; 95% CI, 0.9-2.0; current smokers: OR, 1.7; 95% CI, 0.8-3.5). Furthermore, recent studies from Germany and France also showed similar association of smoking and increase odds of vasculitis (173,180). These findings suggest that smoking is an important confounding factor that should be considered when studying the impact of the environment on the onset of vasculitis.

### 1.6.3. Limitation of current literature and thesis rationale

The past four decades has seen a rise in the number of studies reporting on a range of potential risk factors associated with vasculitis. The most notable findings shows that infection is tightly associated with the onset of AAV, but the exact microorganisms involved are not clear. Smoking also showed dose dependent relationship with both AAV and GCA. Outdoor and indoor airborne exposures through one's occupation are beginning to show adverse association with vasculitis. The role of organic and non-organic particles released during natural disasters including outdoor air pollution exposures have received little attention in vasculitis research. Systematic reviews on airborne exposures and

vasculitis are lacking. For occupation airborne exposures where there is more evidence, there are challenges in comparing the effect estimates from different studies due to large heterogeneity in recruitment, exposure assessment and disease definition. Furthermore, most of studies reporting on environmental risk factor of vasculitis are clinic-based studies, and are limited in power, subjecting them selection bias. Exposures are often derived from self-reported questionnaires and may be prone to recall bias. In order to quantify and understand the environmental footprint associated with airborne exposures, it is essential that we obtain clear measures of exposures, not only at small area and population level but that extend to regional and country level.

The recent rise in data systems including environmental surveillance of air quality and personalised health monitoring provide enormous opportunity to investigate the role of environmental exposure in the onset of vasculitis. One such example is the use of widely collected data on ambient air pollution and meteorology. These data provide quantitative measures on environmental exposures that a person is exposed to throughout their life and includes leading air pollutants that have historically shown to be associated with adverse health event (181). It is important to note that these pollutants cover an array of ambient particles and gases that include dust particles from organic and inorganic source and urban and industrial gases emitted from power stations, traffic, and other industrial related processes. Understanding the role of air pollution in driving the risk associated with vasculitis onset will give insight into who is potential at risk in the general population and can provide mechanistic pathways that lead to the onset of disease.

## 1.7. Thesis aims and structure.

### 1.7.1. Overview of the thesis

The aim of this thesis is to understand the short and long-term impact of outdoor air pollution exposures on the onset vasculitis. The thesis pays particular attention to the role of geography, seasonality, and their effects in explaining the incidence of vasculitis. ANCA-associated vasculitis and giant cell arteritis were used as case studies in the result chapters as they are more common than other small and large vessel vasculitis.

### 1.7.2. Research Questions

#### 1.7.2.1. What is the long-term impact of air pollution exposures on vasculitis onset?

This question was investigated by first undertaking a systematised review of studies reporting on the association between airborne exposures and onset vasculitis. Secondly, an environmental wide association study approach was used to investigate the association between multiple air pollution exposures and the onset of vasculitis in UK Biobank and the Scottish Morbidity Record (SMR01).

.

#### 1.7.2.2. What is the role of geography in mediating the effects of air pollution on the onset of vasculitis?

This question was investigated by conducting a series of multivariable analyses of cross-sectional data from UK Biobank and SMR01 stratified based on population density measures. The effects of air pollution were stratified based on urban and rural geographies, so to investigate for possible geographic disparity in effects of air pollution on vasculitis

#### 1.7.2.3. What is the temporal and seasonal pattern in the onset of vasculitis, and can it provide insight about the possible role of environmental exposures?

This question was investigated by conducting a time series analyses of routine health data of patients diagnosed with AAV between the period of 1996 to 2020 in the SMR01. Several modelling approaches were used to assess for peaks in incidence of AAV and to related them to a possible environmental agent. Seasonal variation was also assessed by comparing the frequency of AAV cases from the different months of the year.

### 1.7.3. Thesis structure

Chapter 2 presents a systematised review of studies reporting on the effect outdoor air pollution and occupational airborne exposures on vasculitis. Studies reporting on the seasonal variation of vasculitis were also assessed. Chapter 3 provides a detailed description of the data resources used to address the research of this thesis. It describes how vasculitis cases were ascertained from the UK Biobank, and the Scottish morbidity record. Chapter 4 will provide a detailed description on the environmental linkage method that were used in this thesis and will give a demonstration on how they can be applied to study environmental risk factors of vasculitis. The results chapters (chapter 5 to 7) will report finding of the analyses showing an association between air pollution, seasonality, and the role of geography in mediating the risk of vasculitis. The last chapter will discuss the implication of thesis findings for clinicians, public health practitioners and policy makers. This will be discussed in context of other literature in the field, acknowledging both the strength and weakness of this thesis and that of others.

# Chapter 2: A systematised review on the impact of air pollution, occupation airborne exposures and seasonality on vasculitis onset.

## 2.1. Overview

Chapter 1 provided an overall summary of the traditional risk factors known to be associated with onset vasculitis. In addition to these risk factors, there are emerging risk factors, such as outdoor air pollution and seasonal exposure, whose impact on vasculitis is not yet recognised. The objective of this chapter is to provide an up-to date systematised evidence of studies reporting on the effects of 1) ambient air pollution exposures, 2) occupational airborne exposures, and 3) seasonality on the risk of AAV and GCA. To achieve this, the review was conducted using a structured protocol based on the Preferred Reporting Items for Systematic review and Meta-Analysis protocol guidelines (PRISMA-P)(182). A clear search strategy detailing the Population Exposure Comparison Outcome (PECO) was employed. The electronic search was conducted in Medline, Embase and several other leading empirical databases, covering publication from 1946 to 2021. The search strings used included many leading air pollutants and occupational airborne exposures known to be associated with adverse health. A pilot search conducted on the 15 September 2021 revealed that there was only one publication reporting the association between air pollution and vasculitis, particularly AAV and none on GCA. Given this, the final search strategy was extended to include other inflammatory autoimmune diseases, specifically rheumatoid arthritis, systemic lupus erythematosus, ankylosing spondylitis, and psoriatic arthritis as outcomes. The goal was to assess if there are any shared effects of air pollution on major inflammatory autoimmune diseases. A summary of this comparison is provided in the discussion section and will inform later discussion in the result section (Chapter 5 and 6)

The search strategy for occupation exposures was targeted at AAV and GCA as outcomes interest. For AAV, the pilot search conducted on September 2021

revealed that there was a systematic review and meta-analysis reporting a significant association between occupation exposure to crystalline silica and an increased odds of AAV (102). This review by Gómez-Puerta and colleagues was published in 2013. It specifically focused on crystalline silica as an occupational airborne exposure but did not assess other major exposures and occupational titles, such as organic solvent or farming, that have been reported to be associated with vasculitis in several observational studies. Additionally, this review's meta-analysis included six articles published between 1993 and 2010. Two of the studies published in 1993 (183) and 1995 (184) were susceptible to hospital control bias and selection bias (185) and generally had inflated odds ratios (OR) and large confidence intervals (Gregorini et al., 1993 - OR 14, 95% CI: 1.7 – 113.8 and Nuyt et al.,1995 – OR: 5.0, 95% CI: 1.4 – 11.6). For both studies, cases (n=16) were matched on 1:2 ratio to hospital controls (n=32). Study participants were recruited at a single clinical centre and the control groups were generally not comparable to AAV population in terms of age and sex. For example, one of the studies focused on males only and the other did not have a sufficient population coverage of the people over the age of 50 years old. To this end, the final search strategy and the reporting on occupation airborne exposures was extended to include both the old (studied published before 2010) and newer studies (studies published between 2010 to 2021). The newer studies included three more case-control studies published between 2015 and 2021 (186–188). These studies had a much larger sample sizes and population coverage and were from different parts of the globe (Europe and Oceania). In line with this chapter's objective, these studies (both the old and new) were pooled in a meta-analysis to provide an up-to-date estimation of the effect of silica on AAV, including a separate meta-analysis on four studies reporting the association between farming, as an occupational title, and AAV. Studies reporting occupation exposure to organic solvents, asbestos, and coal particles and AAV have been synthesised and are narrated in the result section.

Lastly, the search strategy on seasonality and vasculitis included search strings with weather variables (wind, temperature, and humidity). The goal was to assess if there are any studies reporting the association between historic

weather pattern and the onset of vasculitis in adult population. There is a growing body of literature on Kawasaki disease that shows that tropospheric wind pattern may carry dust particles and microbial aerosols from various regions that may serve as potential triggers of vasculitis (189–191). A preliminary run of search strings from the pilot search (09/2021) showed that, to date, there is one study that have reported an association between weather features and the onset of GCA (155). This study found no significant association between weather condition at the time of diagnosis and the onset of GCA. As such, the reporting of seasonality and vasculitis only focused on seasonal variation in the occurrence of AAV and GCA.

## 2.2. Methodology

### 2.2.1. Search strategy and eligibility criteria

As highlighted above, this review was conducted using a structured protocol based on the Preferred Reporting Items for Systematic review and Meta-Analysis guideline for writing a review protocol (PRISMA-P) (182). An electronic search strategy was developed using the Population Exposure Comparison Outcome (PECO) strategy and a data extraction template was generated and validated to allow for a detailed quality assessment of included articles.

#### 2.2.1.1.Population and Control group

The population of interest were adults over the age of 18 years. Outcome was defined as a physician-confirmed diagnosis of AAV or GCA confirmed using one of the following classification criteria: the International Chapel Hill Consensus Conference Nomenclature of Systemic Vasculitides (CHCC), the European Medicine Agency (EMA) algorithm for vasculitis, the American College of Rheumatology (ACR) criteria for vasculitis, or the International Classification of Disease (ICD) criteria in the case of hospital admission record(19,21,22,27,192). An appropriate control group had to share similar demographic or risk profile to patients with vasculitis. This could be individuals who were systematically or randomly recruited from a local hospital or primary

care database and preferably matched by age, sex, and region. Controls sought from the general population or controls with a non or other inflammatory rheumatic diseases were also considered.

### 2.2.1.2.Exposures

### 2.2.1.2.1. Air pollution

The following air pollutants were included in the electronic search strategy: particulate matter with aerodynamic size of less than ten ($PM_{10}$) and particulate matter with aerodynamic size of less than less than two-point five micrometre in size ($PM_{2.5}$), nitrogen oxides ($NO_2$ and NOx), sulphur dioxide ($SO_2$), carbon monoxide (CO) and ozone ($O_3$). These were selected based on existing evidence that show they are important risk factors association with many health outcomes, mainly cardio-respiratory diseases, according to the World Health Organisation (WHO) (193), the European Environmental Agency (EEA) (194), and the United States Environmental Protection Agency (US EPA)(195,196). Some of these pollutants have also shown to be associated with major inflammatory autoimmune diseases, like rheumatoid arthritis and systemic lupus erythematosus (197–199). The search string included in the subject heading (SH) and text words in the title or abstract fields (TW) included broad search term which are summarised below

*[1] SH: Air pollution// or ambient air pollution/or traffic-related pollution/or air pollutants/ or atmospheric pollutant/or particulate matter/or nitrogen oxides/or nitrogen dioxide/ or nitrous oxide/ Sulfur Dioxide/ or Sulphur Dioxide/or Carbon Monoxide/ozone*
*[2] TW: Outdoor pollut*/or air qual*/ or air contamin∗/or atmospheric pollut∗/or atmosphere contaminat∗/or traffic pollut*/or ambient pollut*/ or traffic-related pollut*/or vehicle emi**
*TW: "PM10"/or "PM2.5"/or "NO2"/ or "NOx"/or "SO2/ or "CO"*
*(1) OR (2)*

#### 2.2.1.2.2. Occupational airborne exposures

Articles with data on harmful substances or particles associated with occupational titles listed in the international standard occupational classification system were assessed(200). These exposures are often identified through a job exposure matrix (201,202), or a structured questionnaire validated by an occupational health practitioner or a detailed literature review. Examples of occupational exposures of interest included crystalline silica, asbestos, gas, diesel, dusts, fumes, fibres, and mists. For the search strategy, non-specific search terms such as “Occupational exposures/dusts/chemicals” were included to ensure all potential occupation exposures and pollutants were identified. Below is an example of search strings (subject heading, SH) applied:

*SH: occupational exposures/or occupational dusts/or occupational airborne chemicals/or occupational pollutants/or dust/or silica/or metals/or diesel/or fibres/or asbestos/*

#### 2.2.1.2.3. Seasonality

Studies reporting on the seasonal variation of AAV and GCA were included if they used a meteorological definition of seasonality. Meteorological definition of seasonality is roughly divided into the coldest and hottest quarters of the year. The northern hemisphere hottest seasons are spring (March, April, and May), and summer (June, July, and August), while the coldest seasons are Autumn (September, October, November) and Winter (December, January, February). For studies reported from the southern hemisphere, for example Australia, the seasonal months were expected to be opposite, with the coldest seasons being from March to August and the hottest seasons being from September to February. For studies where no such information was provided, this was assumed to be the case.

*TW: season*/ or seasonal*/ or seasonal temperat*/ or weather condit*/ or wind/or tropospheric wind/ or atmospheric condit*/ or temperature/or outdoor*

*temperat*/ or ambient temperat*/ or air temperat*/or surface temperat*/ or atmosph*)*

### 2.2.2. Data source and study selection

A detailed literature search was conducted on Medline, Embase, Web of Science, Scopus, CINAHL and GreenFILE on October 23, 2021. It included search terms covering studies in English and published between the 1946 and 2021. Additional studies were sought from reference lists of relevant articles. All eligible articles were managed, and duplicates removed, using Mendeley software (v1.19.8). The title and abstract screening were conducted in DistillerSR (DistillerSR v2.35. Evidence Partners, 2021, Accessed Nov 2021 – February 2022 *https://www.evidencepartners.com*. A study was included if it was an original article or a conference abstract with enough data for extraction. Animal studies, case report, case series and studies on children were excluded. A detailed overview of the electronic search strategy can be found in **Appendix 1**. The thesis author performed the search and screening of articles summarised in the result section.

### 2.2.3. Data extraction and quality assessment

A spreadsheet with relevant data fields was used to extract pertinent information from included articles. This was piloted on three studies included in this review and was updated accordingly in the case where relevant information were not being captured. The spreadsheet covered information on study population, methodology and qualitative and quantitative results. Additional information on exposure assessment and time of exposure was included.

All of the included studies were critically appraised by the thesis author using the Joanna Briggs Institute (JBI) critical appraisal checklist for cross sectional and case control studies(203,204). The checklist covered 10 questions on internal validity and risk of bias, paying attention to selection and information bias, confounding, clear reporting, and statistical analysis. These question are: (1) Were groups comparable other than the presence of disease in cases or absence of disease in controls? (2) Were cases and controls matched

appropriately? (3) Were same criteria used for the identification of cases and controls? (4) Was exposure measured in a standard, valid and reliable way? (5) Was exposure measured in the same way for cases and controls? (6) Were confounding factors identified? (7) Were strategies to deal with confounding factors stated? (8) Were outcomes assessed in a standard, valid and reliable way for cases and controls? (9) Was the exposure period of interest long enough to be meaningful? (10) Was appropriate statistical analysis used? The JBI appraisal tool does not provide a formal rating scale on the basis that each question cannot be weighted equally and may not give accurate picture of the quality of each study. Authors are encouraged to state how they derive their rating and to report it in the methodology (203,204). As such this review generated a quality rating that was based on how each study fulfilled the JBI criteria or questions. For example, an article that did not fulfil at least four of the ten questions stated above was rate low (score <4), those fulfilling four to seven questions was rated moderate (score: 4 – 7) and those fulfilling seven to ten question were rated high (score: 7 – 10) – see **Appendix 2 and 3**.

### 2.2.4. Data synthesis

A narrative synthesis of extracted data from studies the effect of air pollution, occupational exposures and seasonality on vasculitis was performed alongside tables and meta-analysis of extracted estimates. These estimates are presented in a forest plot and funnel plot and mostly cover the effects of occupation and occupational exposures on AAV. To address the first objective of this review, studies on air pollution and AAV were summarised in a table with odd ratios and 95% confidence intervals together details the study methodology. Similarly, studies addressing the effects of occupational history and exposures on AAV were pooled in a table with odd ratios and 95% confidence intervals being reported together the method of exposure assessment. Where there were more than three studies reporting the effects of occupation exposures on AAV, those studies were pooled in meta-analysis. For example, a random effects model was implemented using the *metan* package in Stata and estimated the overall-effects of crystalline silica exposure and farming as an occupation title on AAV (205). An in between-study heterogeneity

was assessed by examining the Cochrane Q and $I^2$ statistics (206). A p-value of <0.05 was used to capture a statistically significant heterogeneity of between studies(207). The Funnel Plot was used to identify and estimate the amount of publication bias by using the Egger's test (208). For seasonality, most studies reported absolute incidence or frequencies of AAV or GCA across the different seasons of the year. These estimates were often compared using an ANOVA or pairwise t-test. Results from these studies have been pooled in a table with significant p-values being reported.

## 2.3. Results

### 2.3.1. Study selection and characteristics

**Figure 1** shows a PRISMA flow diagram representing selection criteria of included articles. Overall, the electronic search results identified 5,480 potential studies whose titles were screened for inclusion in the review. Studies not related to the review's objectives were excluded, leaving 375 articles which underwent a detailed abstract screening. As highlighted in search strategy, studies were excluding for detail review if the outcome (diagnosis or onset of vasculitis) and the exposures of interests were not the subject of the study. Articles reporting the link between vasculitis flares and environmental exposures were also excluded. In total, 123 eligible articles were screened for full text review with only 24 articles being included in this review. The exclude articles at this stage (n=99) were reporting on the effect air pollution and occupation airborne exposures on rheumatoid arthritis, systemic lupus erythematosus, ankylosing spondylitis, or psoriatic arthritis (Appendix 6). While these studies were omitted, they were kept for background research and are mentioned in the discussion section of the review.

Of the 24 studies included in this review, 19 were full articles and 5 conferences abstracts. Most studies were from continental Europe (n=11), followed by the United States (n=5), United Kingdom (n=4) and Australasia (n=3) including La-reunion (n=1). The only study on air pollution and AAV was conducted in the UK using two patient cohorts from Nottingham/Derby (2007 – 2013) and Norfolk

(2000 – 2010). This study was a conference abstract and did not have all the descriptive summary data. Further information on the patient cohort were sourced from previously published parent studies using the same cohorts (55,209). Overall, the median age of AAV population in this study was 60 and 70 years old with majority of patients being male (60%) in Nottingham/Derby (33) and female (52%) in Norfolk (34). Similarly, studies reporting the effects of occupational exposures and vasculitis were all on AAV. The median age of AAV population in these studies ranged from 50 to 62 years old with some variation in the proportion of male and female patients. Having a history of an autoimmune disease, being younger than 18 years of age and having missing data on ANCA status or having other forms of vasculitis were part of the exclusion criteria. Two of the seven studies were from the US (210,211), one from New Zealand(187), with the rest being from Europe (Italy, Germany, France) and UK (186,188,212,213). In all these studies, exposure assessment was evaluated using a standardised questionnaire administered either by occupational physician or epidemiologist or via mail. There was no mention of the use of a job exposure matrix, a gold standard in the field occupational medicine. Hogan et al., 2006 reported using the National Institute of Occupational Safety and Health guideline to assess crystalline exposure (211). However, this approach still required a generating an indirect estimation of exposure which was conceived by the study's author and does not reflect an exposure matrix model. Lastly, nine of the fifteen studies on seasonality and vasculitis, were on GCA. In these studies, GCA was diagnosed using the American college of rheumatology criteria for giant cell arteritis and or using histopathological evidence from a temporal artery biopsy (TAB) (214,215). In all these studies, the median age ranges from 70 to 76 years older with increased proportion of patients being female (51% to 79%). In general, patients were excluded from a study if they had a negative TAB or had other conditions with polymyalgia manifestations and in some case if they were younger than 50 years old. **Table 2.1** has provided further detailed summary of characteristics of included studies based on exposure of interest for this review.

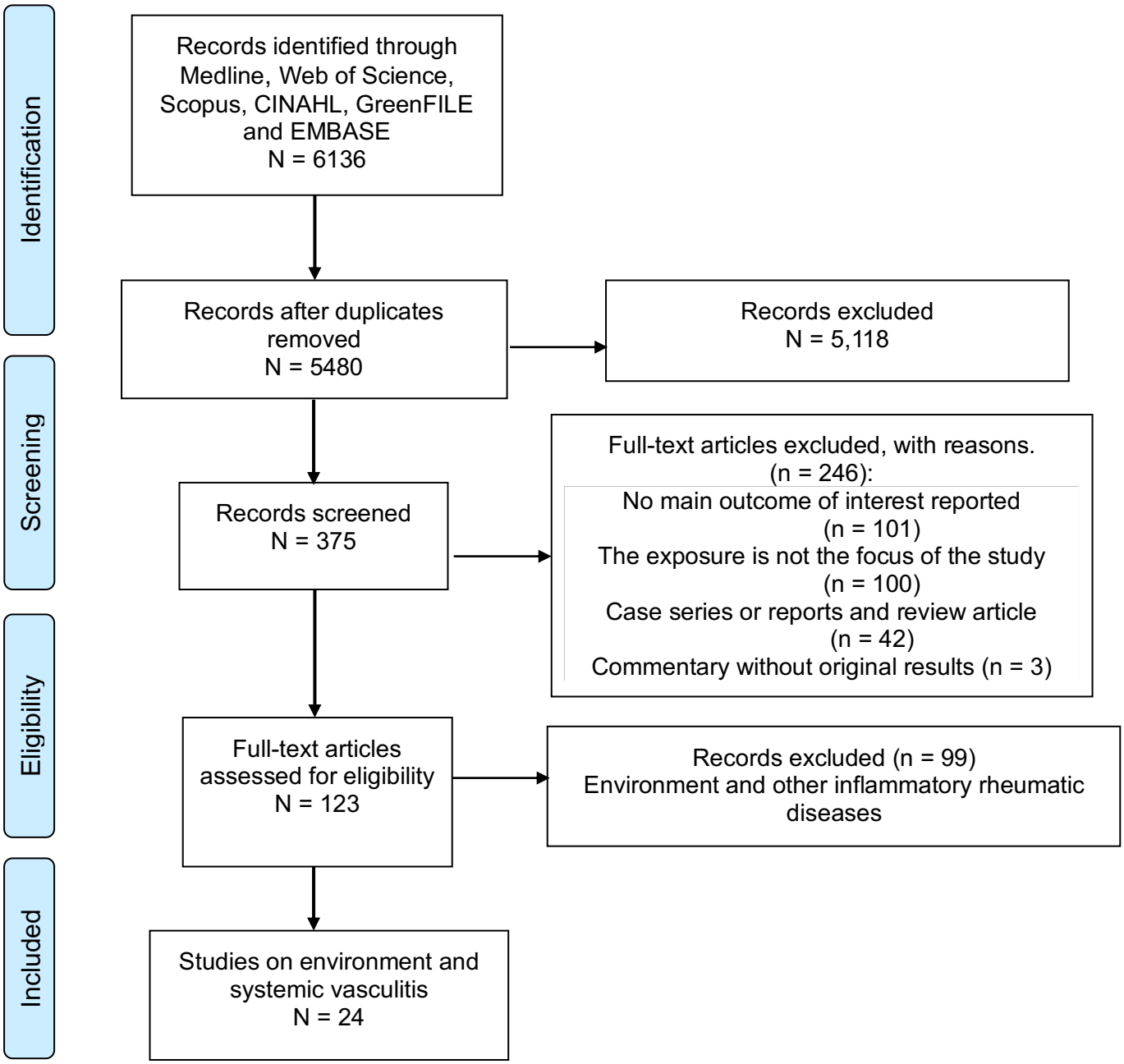



**Figure 2.1:** PRISMA diagram summarising the selection criteria of studies included in this review.

**Table 2.1:** Characteristics of studies on the association between air pollution, occupation airborne exposures and seasonality with AAV and GCA

| Author, year | Country | Study design | Case definition | Age at diagnosis (cases) | Sex (Female) | Criteria for vasculitis diagnosis | Exclusion Criteria | Quality assessment (Joanna Briggs Institute) |
|---|---|---|---|---|---|---|---|---|
| **Air pollution** | | | | | | | | |
| Mahmood-Rao, 2018 (*Conference abstract*) (216) | UK | Case control | AAV | 70.2 | 44% | EMA algorithm for AAV | Not applicable | Moderate |
| **Occupational pollutants** | | | | | | | | |
| Lane, 2003 (213) | UK | Case control | PSV (GPA, MPA, EGPA) | 60.2 | NR† | ACR and CHCC | History of any autoimmune disease, or secondary vasculitis (due to systemic lupus erythematosus infection, cryoglobulinemia or malignancy), death. | High |
| Stamp, 2015 (187) | NZ | Case control | GPA | 53.3 | 57% | ACR, CHCC or EMA | History of other inflammatory diseases | High |
| Hogan, 2006 (211) | US | Case control | ANCA-SVV | 62 | 40% | CHCC | Missing information on age, state of residency | Moderate |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | and ANCA status and not speaking English | |
| Maritati, 2021 (186) | Italy | Case control | EGPA | 53 | 58% | CHCC, ACR | Less than 18 years old at the time of diagnosis, having other vasculitides or cognitive impairment or severe illness that could impede the ability of completing occupation questionnaire | Moderate |
| Willeke, 2015 (188) | Germany | Case control | AAV | 51.2 | 47% | ACR, CHCC, EMA | Not reported | Moderate |
| Beaudreuil, 2005 (212) | France | Case control | AAV | 61.2 | 50% | ACR | Being treated with neomercazole or suffering from infectious disease (particularly endocarditis), antiphospholipid antibodies, or malignancies | Moderate |
| Albert, 2003 (210) | US | Case control | GPA | 57.2 | 50.9% | ACR | Having any other inflammatory, rheumatic, or pulmonary disorder | Moderate |
| | | | | | | | | |
| **Seasonality** | | | | | | | | |
| Chung, 2020 (158) | Australia | Cross sectional | AAV | 63 | 45% | CHCC or ACR | Having EGPA | Moderate |
| Aiyegbusi, 2021 (159) | UK | Cross sectional | AAV | 66 | 54% | CHCC | 16 years and younger, negative ANCA results, EGPA and no kidney biopsy at diagnosis | High |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| González-Gay, 2001 (217) | Spain | Cross sectional | GCA | 68 | 50.90% | ACR | Patients with other diseases that may present with polymyalgia manifestations | High |
| Koldingsnes, 2000 (160) | Norway | Cross sectional | GPA | 48.2 | 38% | ACR | Not reported | High |
| Richier, 2018 (*Conference abstract*) (218) | La-reunion | Cross sectional | GCA | 73.7 | 60% | ACR | Not reported | Moderate |
| González-Gay, 2007 (219) | Spain | Cross sectional | GCA | 75 | 54.50% | ACR | Other diseases that may present with polymyalgia manifestations | Moderate |
| Kisza, 2012 (220) | US | Cross sectional | GCA | 77.3 | 74% | Temporal artery biopsy (TAB) | No TAB results, or no medical record | Moderate |
| Stamatis, 2019 (*Conference abstract*) (221) | Sweden | Cross sectional | GCA | 75.1 | 72% | Not able to assess | Negative TAB | High |
| Khelgi, 2017 (*Conference abstract*) | Australia | Cross sectional | AAV | 66 | 62% | Not able to assess | Not reported | High |
| Gokoffski, 2019 (222) | US | Cross sectional | GCA | 76.4 | 72% | ACR | Diagnoses outside the recruiting centre (Davis medical centre), no TAB prior to 1995 | Moderate |
| Konig, 2021 (155) | Denmark | Cross sectional | GCA | 75.6 | 66.3% | Temporal artery biopsy (TAB) | 50 years and younger, no representative biopsy and no suspicion of polyarteritis nodosa (PAN) based on histopathological criteria | High |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Raheel, 2018 (*Conference abstract*) (156) | US | Cross sectional | GCA | 75.6 | 79% | ACR | Not able to assess | High |
| Mahr, 2006 (223) | France | Cross sectional | GPA | 54.3 | 63% | ACR | No clear GPA by ACR classification criteria and no clear record of time of GPA onset | Moderate |
| Draibe, 2018 (130) | Spain | Cross sectional | AAV | 65.3 | 50.8% | CHCC | No precise month of AAV onset that could be calculated | High |
| Frieta-Gilchrist, 2020 (*Conference abstract*) (157) | UK | Cross sectional | AAV | 65.6 | 53% | Biopsy proven AAV | Not reported | High |
| Petursdottir, 1999 (224) | Sweden | Cross sectional | GCA | >50 | NR | Temporal artery biopsy (TAB) | No TAB | High |

† Not reported

### 2.3.2. Representativeness of included studies.

The two cohorts assessed for the study on air pollution and AAV were generally older for a vasculitis population in the UK, with the combine mean age of 66 years old. The Nottingham-Derby cohort was on average older (median:70.2 years old) compared with the Norfolk patients (60.1 years old). This may reflect true differences in the aging population between these two counties. Regardless of the age differences, the overall incidence of AAV in each cohort was representative of the UK AAV estimates (15.1 to 23.1 cases per million population) seen in chapter 1, with Nottingham-Derby cohort showing an incidence of 23.1cases per million population and Norfolk being 22.6 cases per million population. The male to female ratio was also similar to what is seen in other studies in the UK and globally (51). Additionally, the control group and denominator used covered the general population of the counties as the study used census population data from the office of national statistics (ONS) 2011, including the middle layer super output areas (MSOA) for each of the case-catchment area.

In the case of occupation exposures and AAV, most studies were generally representative of the AAV population in terms of age and sex. The study from Germany, Italy and New Zealand had a generally younger population with a mean age that ranged from 51 and 53 years old(186–188). The rest of the studies had AAV population with a mean age starting from 60 and above which is generally expect. In all the studies, the cases were recruited from hospital databases or specialised clinics (e.g.: nephrology and rheumatology). Hospital Controls were recruited in similar manner in five of the seven studies and included individuals with a non-inflammatory rheumatic disease like gout, osteoarthritis, cardiovascular disease and diabetes or other miscellaneous diagnoses(187,188,210,212,213). Controls were matched by age, sex, and geography in some cases. Two studies recruited from the general population via random sampling involving calling and inviting local residents from nearby regions or extended family members to participate in a study (186,211).

Lastly for studies reporting on the seasonal variation in vasculitis, particularly those focusing on GCA had a representative population with mean age of 70 years or old in majority of studies. The three studies on AAV were also generalisable to the AAV population, except the study by Koldingsnes et al, 2000 in Norway, which had a younger cohort (mean age: 48 years old). Most studies required a positive temporal artery biopsy confirmation and exclusion patients with polymyalgia or a history of autoimmune disease, minimising observation bias and overrepresentation of other rare disease associated with GCA.

### 2.3.3. Appraisal of study methodology and reporting

Ratings of the methodology and reporting bias were assessed using the JBI critical appraisal tool for case control and cross-sectional studies and are summarised in Appendix 4 and 5**.** Overall, the rating for the one study on air pollution and AAV was considered moderate. Reason being there was a limited details in the methodology on the approach used for environmental linkage and timing of exposure. For example, it was not clear whether the period of exposure was the index year (i.e.: the year of diagnosis) or the years preceding the start of disease. For occupational exposures, most studies varied in quality and reporting, and ranged from moderate to high. Studies rated to be moderate were due to limitations in methodology particularly related to the exposure assessment and reporting. Few of the included studies used validated questionnaires to capture lifetime job histories. These were developed by lead investigators together with occupation health specialists and epidemiologist. Irrespective of this, it was not clear whether common tools such as the job exposure matrix or the international occupation classification system were used for the exposure assessment (200,202). These tools are considered gold standard and have been shown to provide quantitative estimates of both the short and long-term impact of occupational exposure such as silica or asbestos on various health outcomes (176,225–228). Lastly, studies on seasonality used historic medical records to assess for seasonal and temporal variation in incidence of AAV and GCA. Half of the studies were rated moderate due limited reporting on case ascertainment or any further exclusion criteria.

### 2.3.4. Finding of the electronic search

#### 2.3.4.1.Effects of air pollution on AAV

As highlighted earlier, the results from the electronic search on air pollution and vasculitis identified one study on AAV, a conference abstract published from the British Society of Rheumatology (Table 2.2). The study assessed the short-term impact of residential air pollution at the lower layer super output area (LSOA) and its effect in ANCA-associated vasculitis. The study consisted of two cohorts from Nottingham-Derby (2007-2013) and Norfolk (2000-2010) area. The Nottingham-Derby cohort had 107 cases with census population controls of 741,071, while the patients from the Norfolk cohort were 86 with census population controls of 459,000. The standardise incidence was 14.44 cases per hundred thousand population in Nottingham-Derby and 18.74 case per hundred thousand population in Norfolk. The study utilised air quality index data from the office national statistics (ONS) for England for 2008 and covering the 'study period', although this period is not clearly defined in the method, in terms of how far back or forward for the linkage done. The study focused on $PM_{10}$, $SO_2$, and $NO_2$ and it assigned pollution exposure by allocating decile exposure measures to each case and a complete decile distribution for the whole geographical area in each cohort. Differences in exposure between case and controls were compared using chi-squared test, and logistic regression. Overall, cases from Nottingham/Derby showed to be from areas with low levels of air pollution, particularly $SO_2$ and $NO_2$ compared with the Nottingham/Derby general population. This was statistically significant when assessed for combined indices ($SO_2$ odds ratio: 0.84, 95% CI: 0.74 - 0.95, $NO_2$ odds ratio: 0.85, 95% CI: 0.75 – 0.96). For the Norfolk area, the effects of air pollution were around the null and therefore were not statistically significant ($PM_{10}$ odds ratio: 0.99, 95%CI: 0.78 - 1.25, $SO_2$: 1.01, 95%CI: 0.86 - 1.20, $NO_2$: 1.02, 95%CI: 0.89 - 1.18) (216). This variation in effects of air pollution by geography may reflect the heterogeneity in air pollution distribution across the two counties. Stratified analysis based on urban and rural classifications might have given some indication about the effects of air pollution across space and geography

and within each cohort. This limitation will be address in later chapters to gain insight the varying effects of air pollution across urban and rural geographies.

**Table 2.2:** Summary of one study reporting on the association between air pollution exposures and ANCA-associated vasculitis

| Author (year) | Case definition | Control definition | Sample size | | Matching/ Adjusted | Type of air pollutants | Assessment of air pollution exposure | Time of exposure | Odds ratio (95% CI) |
|---|---|---|---|---|---|---|---|---|---|
| | | | Case n | Controls n | | | | | |
| Mahmood-Rao, 2018 (216) | AAV | General population controls (Census, 2011) | 107 and 86 | 741,071 and 459,000 | age, sex | $PM_{10}$, $SO_2$, $NO_2$ | Modelled pollution data incl. ground monitoring stations and satellite data | At the year of diagnosis | **Nottingham/Derby**<br>PM10: 0.85 (95% CI: 0.72 – 1.00)<br>SO2: 0.84 (95% CI: 0.74 – 0.95)<br>NO2: 0.85 (95% CI: 0.75 – 0.96)<br><br>**Norfolk**<br>PM10: 0.99 (95%CI: 0.78 - 1.25),<br>SO2: 1.01 (95%CI: 0.86 - 1.20),<br>NO2: 1.02 (95%CI: 0.89 - 1.18) |

### 2.3.4.2.Effects of occupational airborne exposures on AAV

Table 2.3 provides a detailed description of studies reporting the association between occupational exposures and AAV. While most studies aimed to report on all major airborne chemical exposures, such as mineral dust, fumes, diesel, and biological dust, only crystalline silica exposure and farming as an occupation title were the two factors mostly reported on in over half of the included studies (187,188,211,213). The rest of reported exposures included volatile organic compounds, metals, pesticides, inorganic dust, and organic solvents (186,210,212). Of those reporting solely on silica and farming, the heterogeneity between studies was small and not statistically significant ($I^2$=0.0%, p=0.458 and $I^2$=0.0%, p=0.548, respectively). The variation seen was due to chance and is not reflective of real differences between studies - Figure 2.3 and 2.5. For silica, the pooled estimate of its effect on AAV was nearly two-folds (pooled odds ratio: 1.95, 95% CI: 1.41 – 2.71) - Figure 2.2. Similarly, lifetime exposure to farming, as an occupation, was associated with 80% (pooled odds ratio: 1.80, 95% CI: 1.33 – 2.46) increased odds of AAV - Figure 2.4. When assessing this effect by disease subtype, individuals with EGPA were at higher risk compared with MPA and GPA in two studies focusing farming history as main exposure of interest (186,188). Other major exposures reported in the included studies are organic solvents, asbestos, and coal. For organic solvents, lifetime exposure was associated with over two-fold (odds ratio: 2.20, 95% CI: 1.14 – 4.25) increased odds of EGPA in Italy (186). Occupational exposure to asbestos was also reported to be associated with 30% increased odds of AAV (odds ratio:1.3, 95% CI: 0.5-3.2) in France. Lastly, long-term exposure to coal (through mining, electricity generation, steel making and coal furnaces) was not significantly associated with AAV (odds ratio: 1.80, 95% CI: 0.674-4.875) when compared with hospital controls with osteoarthritis and gout.

**Table 2.3:** Studies on occupational exposures and ANCA-associated vasculitis

| Author, year | Case definition | Control definition | Sample size | | Matching/ Adjusted | Occupation exposures | Time of exposure | odds ratio (95%CI) |
|---|---|---|---|---|---|---|---|---|
| | | | Cases (n) | Controls (n) | | | | |
| **Lane, 2003** (213) | PSV (GPA, MPA, EGPA) | Hospital controls (hospital inpatients and outpatients with NIRD*) | 103 | 220 | Age | Farming and Crystalline silica | Any time during working life | Farming - 2.2 (1.2–3.8) Silica - 1.4 (0.7–2.7) |
| **Stamp, 2015** (187) | GPA | Non-inflammatory musculoskeletal disease (osteoarthritis or fracture) | 49 | 196 | Age, sex | All jobs, farming, and Crystalline silica | > 6 months prior to index date | Farming 1.7 (0.8–3.4) Silica 3.1 (1.3–7.4) |
| **Hogan, 2006** (211) | ANCA-SVV | General population | 129 | 109 | Age, sex, education, and region | Crystalline Silica | Lifetime job history held for 12 months or longer | Silica 1.6 (0.9 to 2.8) |
| **Maritati, 2021** (186) | EGPA | General population | 111 | 333 | age, sex, region | Farming, organic solvent, Crystalline Silica | Lifetime job history | Farming 2.10 (1.19–3.73) Silica 2.26 (1.10–4.62), Organic solvents - 2.20 (1.14–4.25) |
| **Willeke, 2015** (188) | AAV | RA and Large vessel vasculitis or Polymyalgia Rheumatica (PMR) | 189 | 190 | age, sex | Farming | Ever worked on a farm | Farming - 1.24 (0.66–2.32) |
| **Beaudreui, 2005** (212) | AAV | Hospital Controls (In-patients with diabetes, CVD, or | 60 | 120 | age, sex | Silica and asbestos | Lifetime job history held more than 6 months | Silica - 3.4 (1.1–9.9), Asbestos 1.3 (0.5-3.2) |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | other miscellaneous diagnosis) | | | | | (median =16.3 years) | |
| **Albert, 2003**(210) | GPA | Hospital controls (Osteoarthritis and gout) | 53 | 103 | age, sex | inhalants, liquid solvents, solid metals, and other chemicals associated with common hobbies and industries | Lifetime job history held more than 6 months | Silica - 1.47 (0.55 - 3.93), Coal - 1.80 (0.674-4.875) |

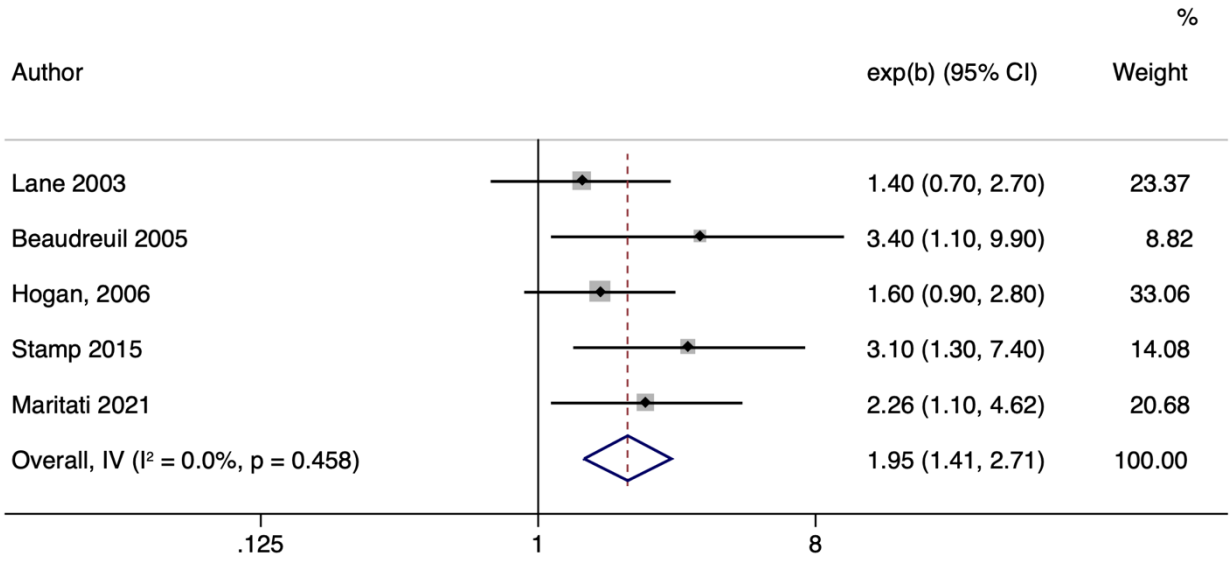



**Figure 2.2:** Meta-analysis on the association between Crystalline silica exposure and ANCA-Associated vasculitis

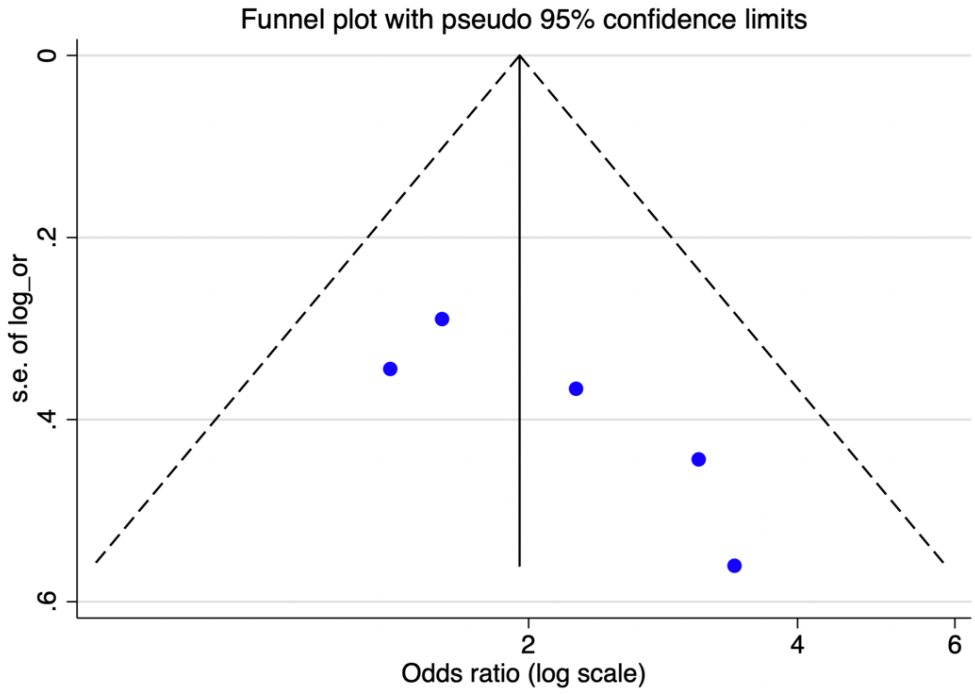



**Figure 2.3:** Funnel plot showing the publication bias of studies on the association between crystalline silica exposure and ANCA-Associated vasculitis.

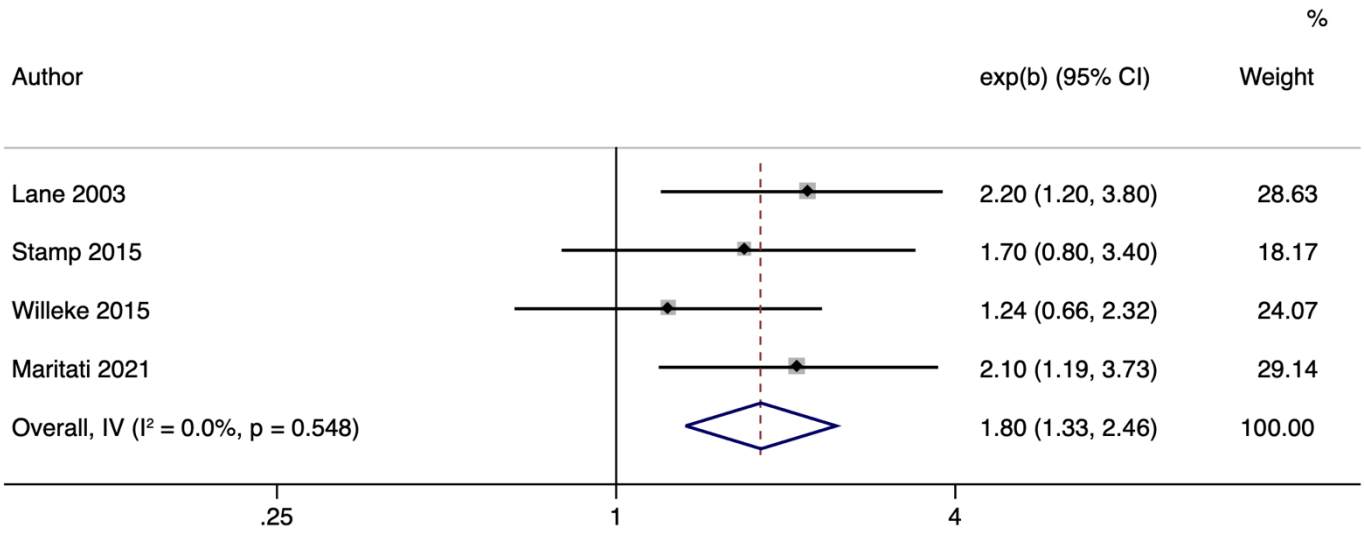



**Figure 2.4:** Meta-analysis on the association between farming occupation and ANCA-Associated vasculitis.

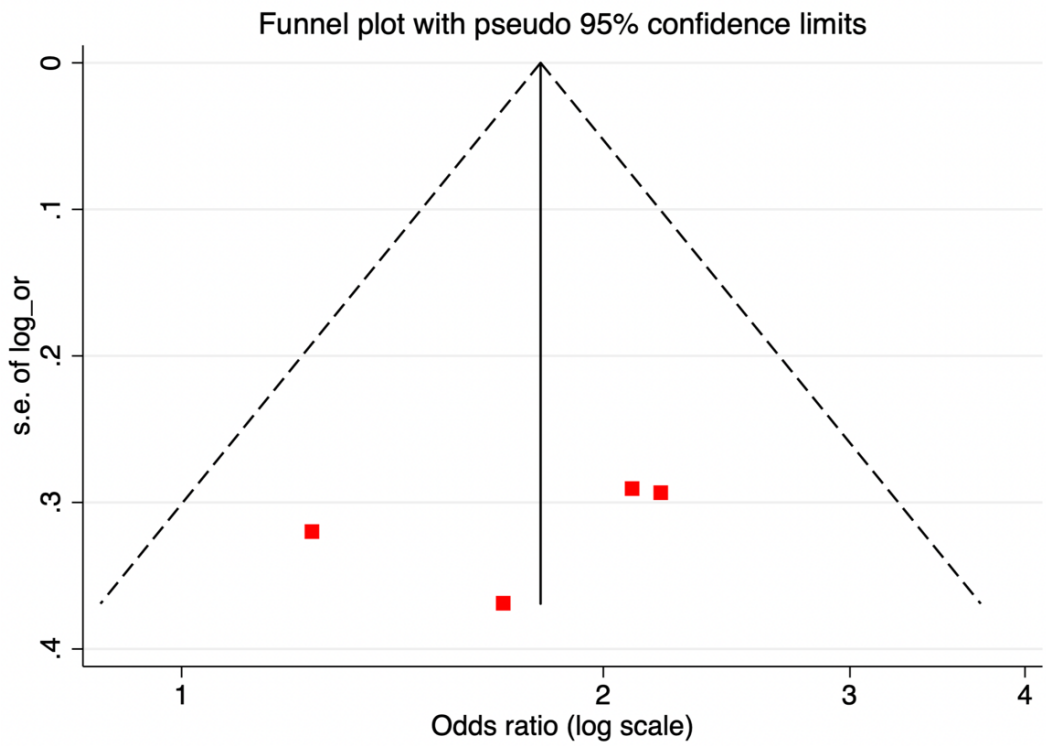



**Figure 2.5:** Funnel plot on the association between lifetime exposure to farming environment and ANCA-Associated vasculitis risk.

### 2.3.4.3.Seasonal variation in vasculitis

The results on the seasonal variation of vasculitis revealed a mixed pattern in the annual occurrences of AAV and GCA. A total of seven studies reporting on the seasonality of AAV were identified. Five studies used the date of diagnosis as proxy measure of disease onset while two studies reported using the date of first symptom onset. Several methodologies were used in to quantify seasonal variation in disease onset. These varied from hypothesis testing approach using p-values from descriptive statistics (Chi-Square and one-way ANOVA) comparing the frequency or mean of AAV cases from different seasons of the year to regression-based methodology (Poisson regression) using seasonality as a predictor in a model adjusting mostly for age and sex. Overall, over half of the studies (57%, n=4) reported no seasonal variation in AAV, regardless of the method used. Two studies reported a higher incidence of AAV in warmer seasons (Spring and summer, p-value <0.01). A study by Draibe et al.,2018 showed an increase in incidence of AAV during winter, when using the date of first symptom onset as a measure of disease onset. This variation in results between the last three studies may reflect variation in the methodology used to define disease onset. For GCA, this variation in case definition was not seen as onset of disease was defined using the date of diagnosis after a confirmed temporal artery biopsy. Additionally, seasonal variation was assessed using a regression-based method (logistic or Poisson regression) in half of studies, the rest used chi-square and ANOVA to compare seasonal frequencies in occurrence of AAV. The key findings from assessing studies reporting on seasonality of GCA shows that incidence of GCA may be increased during the warmer months of the year, spring, and summer. Half of the studies reported a statistically significant increase in number of GCA cases during the months of April to July ($p<0.05$). Three studies reported no seasonal variation and one study higher incidence of GCA cases during winter and autumn months ($p<0.041$).

**Table 2.4:** Studies reporting on seasonal variation of AAV and GCA

| Author, year | Study years | Study period | Cases | Definition for disease onset | Statistical analysis | Sample size | Findings |
|---|---|---|---|---|---|---|---|
| **ANCA-Associated Vasculitis** | | | | | | | |
| Chung, 2020 (158) | 2002-2017 | 16 years | AAV | Date of diagnosis | Man-Whitney and Holm-Bonferroni correction | 100 | No seasonal variation (p=0.08) |
| Koldingsnes, 2000 (160) | 1984-1998 | 15 years | GPA | Date of first symptom onset | Poisson regression | 55 | No seasonal variation (p=NS) |
| Aiyegbusi, 2021 (159) | 2014-2018 | 5 years | AAV | Date of biopsy-proven diagnosis | One way ANOVA | 339 | No seasonal variation (p=NS) |
| Mahr, 2006 | 2001-2004 | 4 years | GPA | Date of diagnosis | One-way goodness of fit Chi Square | 59 | Higher incidence of **GPA** in **summer** (p=0.001) |
| Draibe, 2018 (130) | 2001-2014 | 15 years | AAV | Date of first of symptom onset | One-way goodness of fit Chi Square | 234 | Higher incidence of **AAV** in **winter** (p=0.003) |
| Khelgi, 2017 (Conference abstract) (229) | 2006-2016 | 13 years | AAV | Date of biopsy-proven diagnosis | Not reported | 29 | Higher incidence of **AAV** in **Spring** (p=0.001) |
| Frieta-Gilchrist, 2020 (Conference abstract) (157) | 2014-2018 | 5 years | AAV | Date of biopsy-proven diagnosis | One-way ANOVA | 339 | No seasonal variation of AAV (p=NS) |
| **Giant Cell Arteritis** | | | | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Richier, 2018 (Conference abstract) | 2005-2017 | 13 years | GCA | Date of diagnosis | Not reported | 60 | No seasonal variation (p=NS) |
| González-Gay, 2007 (219) | 1981-2000 | 19 years | GCA | Date of diagnosis and symptom onset | Chi-square test | 255 | No seasonal variation (p=0.13) |
| Kisza, 2012 (220) | 1994-2011 | 8 years | GCA | Date of diagnosis | Chi-square test | 215 | No seasonal variation (p=0.40) |
| Stamatis, 2019 (Conference abstract) (221) | 1997-2016 | 10 years | GCA | Date of diagnosis | One way ANOVA | 1202 | Higher incidence of **GCA** in **spring** and **summer** (p=0.04) |
| Gokoffski, 2019 (222) | 2003-2014 | 12 years | GCA | Date of diagnosis | Logistic regression | 29 | Higher incidence of **GCA** in **summer** (p=0.028) |
| Konig, 2021 (155) | 1990-2018 | 19 years | GCA | Date of confirmed temporal artery biopsy (TAB) | Logistic and Poisson regression | 336 | Higher incidence of **GCA** in **summer** (p=0.037) |
| Raheel, 2018 (Conference abstract) (217) | 1995-2009 | 15 years | GCA | Date of diagnosis | Quasi-Poisson regression | 248 | Higher incidence of **GCA** in **summer** (p=0.018) |
| Petursdottir, 1999 (217) | 1976-1995 | 20 years | GCA | Date of confirmed TAB | Poisson regression | 665 | Higher incidence of **GCA** in the late **winter** and **autumn** (p=0.041) |

## 2.4. Discussion

### 2.4.1. Summary of findings

The first two aim of this systematised review were set to identify and provide up-to-date effect of air pollution exposure and airborne exposures related to occupation titles and their association with vasculitis. 9 articles were identified with data on air pollution (n=1) and occupational airborne exposure (n=8) and AAV. Results from the electronic search on air pollution indicated a clear need for more studies looking at the impact of air pollution on the onset of vasculitis, specifically AAV and GCA. The only study identified from the UK showed that the effects of air pollution varied based on geography, with patients from the Nottingham/Derby being residents of areas with historically low levels of $SO_2$ and $NO_2$ compared with their corresponding background population. These effects were not seen in the second cohort used from the Norfolk areas. The effects of $PM_{10}$, $SO_2$ and $NO_2$ were not statistically significant. For studies reporting on the relationship between occupational exposures, like silica, asbestos, organic solvents, and farming, and AAV, showed that indeed exposure to silica through jobs held for more than 6 months were associated with increased likelihood of developing AAV. Similarly, working on a farm for more than six month or throughout life was associated with close to 2-fold increased risk of AAV. Occupational airborne exposure like asbestos and organic solvents were also associated with more than 30% increased risk for AAV. Furthermore, studies reporting on the seasonal variation of AAV and GCA provided some evidence to indicate that the incidence rate of GCA may be higher in warmer season (spring and summer) while no clear seasonal pattern were seen for AAV.

### 2.4.2. Strengths and limitation of included studies

The first objective of this review aimed to identify all possible evidence that exist on the effects of air pollution exposure on vasculitis onset. The only study identified on this provided important evidence that showed that the effects of air

pollution were heterogeneous based on geography. This finding is important as it informs future studies on careful interpretation of pollution results when the effect of geography is not considered. Additionally, the study is generalisable to the AAV population and has provided the first data on the potential short-term impact of air pollution on vasculitis. Key limitation of the study was related to the assignment and reporting of the exposure estimates. It was not clear from the reporting that the air pollution estimates were based on cumulative historic exposure based on preceding months to years before the start of the disease. The use of air quality indices is also not very common practice in air pollution studies. Index measure may not accurately capture individual exposure needed for case control studies comparison for a rare disease like AAV.

The second objective of this review provided the first up-to-date meta-analysis of the association between occupation airborne exposures, in particular silica, while minimising the inflated effects from older studies (1993 and 1995) included in the previous meta-analysis(102), which was subject to selection and hospital control bias (185,230). One of the key strengths of many of the included studies especially those published after 2010, was the general rise in sample size and power to detect statistically meaningful effects that could be attributed to silica or asbestos or coal particle exposure(186–188). Majority of these studies were representative of the AAV population, and all of them used the ACR and CHCC classification criteria with some including ANCA testing to confirm a diagnosis of AAV. The key limitation is that none of the studies used a job exposure matrix (JEM) to ascertain exposure dose associated with different occupation titles. Most studies used a validated, and an expert led questionnaire which varied between studies and are not standardised to capture silica unlike a JEM which is often designed to capture other major exposures, such as fine particles or gases associated with multiple health outcomes(231).

Lastly, the third objective of this review showed that nearly half of the studies included for the seasonal variation of vasculitis were limited to single care centres with limited spatial coverage and power to detect significant temporal signal associated with seasonality (232). The one study from Sweden that used

administrative health data with more than 1000 patients with GCA was among few exemplar studies to capture a clear and significant temporal and seasonal variation in GCA onset (221). Furthermore, the heterogeneity in between study findings may have been impacted by the various in definition of disease onset. As highlighted earlier, some studies reported using the date of first symptom onset or date of biopsy-proven diagnosis or clinically confirmed date of diagnosis that preceded ANCA testing. The seasonal variability capture in some studies may have only represented the different stage of disease. For example, one of the few studies reporting on the seasonality of AAV (Draibe et al.,2018) and that used the date of first symptom as a measure of disease onset reported higher incidence of AAV in winter(130). This increase in winter symptom onset could be explained by environmental exposures associated with seasonality. Such exposures range from influenza to air pollution associated with winter heating and energy consumption to low UVB exposure and potentially lower vitamin D. The increase of AAV cases reported in spring and summer from studies using a date of diagnosis may represent similar lagged effects of seasonal exposure mentioned and may be impacted by diagnostic delays seen from the time of admission to diagnosis .This diagnostic delay can range from 0 to 53 days (233). Regardless of the reported limitations, the results from these studies provided a wider picture about the role of seasonality in explaining the temporal variation in the occurrence of AAV. For GCA, it can be seen there is a trend towards increased incidence association with warmer seasons, similarly giving indication that there may be lagged seasonal effect associated with environmental trigger associated with winter and autumn.

### 2.4.3. Limitation of this review

This review was conducted using recommended practice for screening and selecting studies for a systematised review. This included proper use of data extraction and critical appraisal tools which allowed for appropriate assessment of bias of each included studies. Overall, studies were screened for eligibility based on information provided in the abstract and it is possible that necessary data needed to address the review's objective could have been missed. This is particularly true if environmental exposures assessed here were not the main

subject of the study but were included in the analysis and were highlighted in result. It is sometime common that studies reporting on the incidence and prevalence of vasculitis might look at seasonality but not explicitly report on it. This review also included conference abstracts which meant that additional information could not be extracted for some of the studies. In few cases, the necessary information could be extracted from the parent study but not all. Additionally, the JBI critical appraisal tool used for the seasonality studies was set up for cross sectional studies and did not capture time series component related to power and time-series coverage. These components were added to the appraisal tool to capture them and for rigorously assess of each the included study. Lastly, another limitation is the screening and selection of studies were conducted by the thesis author only and were not further checked and compared by another reviewer.

## 2.5. Chapter summary and next steps

The past three decade has seen a rise in number studies showing adverse effects of occupation airborne exposures on vasculitis. Fine particles such as silica and asbestos have stood out to be important risk factors. Alongside this, emerging evidence has recognised that certain occupations titles such as farming, and their associated environment (e.g.: rurality) may also impact the risk for AAV. This review has provided an up-to-date summary of studies that compliments this evidence while broadening the scope to include the role of outdoor air pollution in the onset of vasculitis. A clear gap was identified especially with the lack of evidence on the role of air pollution and their association with AAV or GCA. An important lesson learned from the only study air pollution and AAV shows the importance of considering geography when investigating the impact of airborne exposures on vasculitis. Such lessons will be used to inform analysis in the result chapters. The next two chapters will first provide a detail summary of the data sources and recruitment approach of the study population used this thesis. They will also discuss the methodology of environmental exposure research. Highlighting the strengths and limitation of each method in context history and case studies using those methods. The method section will also provide a demonstration by means of a pilot study of

how these methods can be used to address environmental aetiologies of vasculitis.

# Chapter 3: Description of UK Biobank, UKIVAS and Scottish Morbidity Record resources

## 3.1. Overview

This chapter will provide a detailed description of the cohort composition, outcome definition and selection processes from which vasculitis patients were derived and sourced used to investigate the relationship between airborne risk factors and vasculitis in United Kingdom. The chapter includes a summary of three main data resourced used in this these, 1)the UK Biobank (UKB), 2)UK Ireland vasculitis registry (UKIVAS) and 3) Scottish Morbidity Record (SMR01). All three resources provided an excellent platform that allowed for adequate identification of patients living with vasculitis in UK and included health-related data that covered vasculitis patients diagnosed over the past two decades. The data sources were also diverse in terms of demographic and geographic make-up, with the UKB providing additional exposures that were used to enhance our understand of the role of the environment on vasculitis. For UKB and SMR 01, participants with vasculitis were identified using the International Classification of Disease (ICD) chapter codes for systemic vasculitis, and specifically AAV and GCA. For UKIVAS, patients with vasculitis were diagnosed by specialist physicians using the Chapel hill Consensus Conference nomenclature. All reported studies received favourable ethical approval from the National Health Service (NHS) Research ethics committee (UKB: REC 21/NW/0157 and UKIVAS:REC 10/H1102/77 – Appendix 7). Access to the Scottish vasculitis cohort was achieved under the VOICES study which also received favourable ethics approval from the NHS Scotland Public Benefit and Privacy panel for Health and Social care (PBPP 1819-0069 – Appendix 8-9). Figure 1 provides a detailed overview of how these data sources were used and set up to address the overall research objectives of this thesis

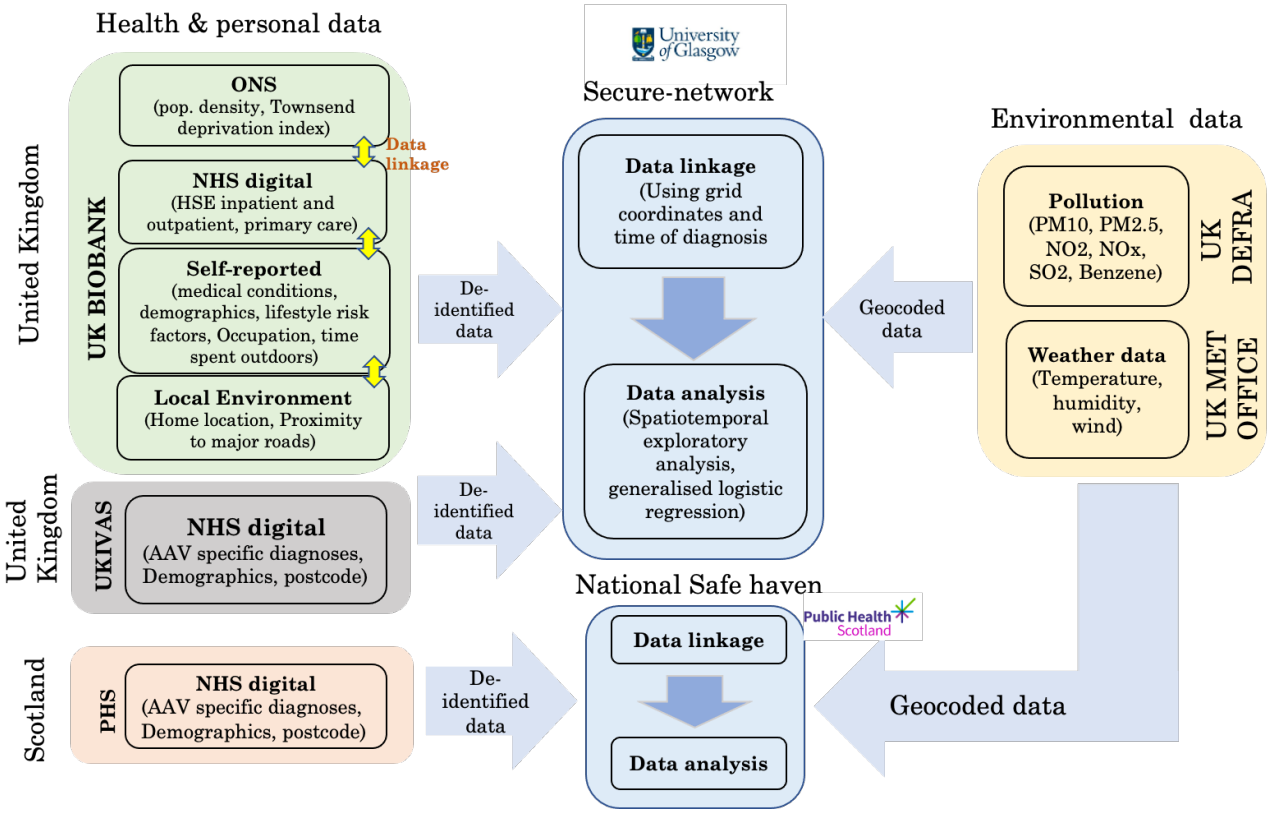



**Figure 3.1**: An overview of the resources used in this thesis to investigate the relationship between environmental exposures and vasculitis.

## 3.2. Description of UK Biobank cohort

### 3.2.1. Background, study design and objectives

UK Biobank is large population-based cohort study (n=502,459, as of February 2021), established to investigate the genetic and non-genetic causes of poor health and disease in middle- and older-aged people (234). UKB was established by the Medical Research Council and the Wellcome trust and have to date collect a wide range of exposures through detailed questionnaires and physical measures, as well as biological sampling to allowed for important assays to be performed for genetic, proteomic, metabolomic and biochemical characterisation (235). UKB aims to combine extensive precise assessment of exposure with comprehensive follow up and characterisations of many different health-related outcomes, as well promote innovative science by maximising open resource access. Recruitment of more than 500,000 participants

consenting to be followed and have their health outcomes data linked from the NHS took place between 2006 and 2010. The large sample size was based on statistical power calculations for nested case-control studies (236) that show that 5,000 to 10,000 cases of any particular conditions would be required to reliably detect odds ratios for the main effects of different exposures of 1.3 to 1.5 (for genomic studies) and 20,000 for detection of interactions with ORs of at least 2.0. The selection of the age range (40 - 69 years) represented a pragmatic approach between participants being old enough for sufficient incident health outcomes during the early years of follow up and young enough for the initial assessment to occur before the start of disease that could be attributed to important measured exposures. UKB provides its data and sample as wide as possible for health-related research in the public interest by all bona fide researchers, from academic, charity, public, and commercial sectors. This access is both for UK and international researchers, without preferential or exclusive access for any user. UKB provides an online access process, which aims to be fair, transparent, and streamlined (234).

Applications are only approved for researchers that are in the public interest and the data required or will become available. Only de-identified data are provided to researchers, who must sign a material transfer agreement, undertaking not to attempt to identify any participant, to keep the data secured, and to use it only for the purposes of the approved research. UKB board of directors has overall responsibility for its direction and management. An executive management team, with epidemiology, clinical, management, laboratory, legal and communication expertise, oversees the development and day to day management of the resources and is responsible for staff at coordinating centres. This governance structures have facilitated effective working between scientific and management disciplines, allowing UK Biobank to respond to advice from wider network of researchers on the most scientifically valuable design and development of resources, with project management and implementation being the responsibility of the UK Biobanks executive management team and dedicated staff.


**Figure 3.2**: Spatial location of UK Biobank recruitment centres and their sample size, Sourced from Sarkar et al.,2015, Annals of GIS (237)

### 3.2.2. Cohort recruitment and baseline assessment

Due to the broad scope of the UKB resource, the emphasis of the baseline assessment was concentrated on the known and potential risk factors for outcomes already known or are projected to become important public health concerns for the adult population. Each question was set with respect to feasibility, comprehension, and acceptability, as well as time taken to complete. The response distribution was examined in pilot studies which aided the final selection and presentation of suitable questions. The UK Biobank questionnaire was administered in two sequential parts during the assessment centre visit: a

touchscreen self-completed questionnaire followed by a computer-assisted personal interview (CAPI). Due to the relative staff costs for self-completed versus interviewer-administered questions, topic areas and questions considered of an exploratory nature haven been restricted to the self-completed questionnaire (wherever possible), and questions that needed to be asked by an interviewer required greater evidence of their value to be included. Repeat assessment of certain lifestyle factors (e.g.: diet) was deemed important for a subset of the UKB population to account for their variation over time.

To optimise on the accuracy and completeness of the data collected while also maximising the efficiency of the process of data capturing, the UKB devised computer technology to record questionnaire responses based on platforms used previously in large scale studies. This technology has been piloted to determine its usability and acceptability among potential participants and has been enhanced in the light of that experience. The self-administered touchscreen questionnaire was used to collect most of the information, taking 30 minutes to complete and single member of staff able to monitor and assist about 10-12 participants simultaneously. Information not collected via touchscreen system were subsequently later collected via CAPI which was designed to last for 5-10minute to control staff costs. A pre-visit aide memoire was provided to participants prior to attending the assessment centre so that they can note certain information (e.g.: medications, family history and birth details) that may be difficult or time-consuming for them to recall during the visit.

Pre-coded lists of diseases, drugs, and occupations are built into the CAPI system, along with structured search facilities, to help this information to be recorded (and automatically coded) both rapidly and completely. Use of inbuilt cross-checks between relevant questionnaire responses and check messages when extreme values were entered or when no value were entered or when no value was provided, were part of the general approach to improve data quality and efficiency.

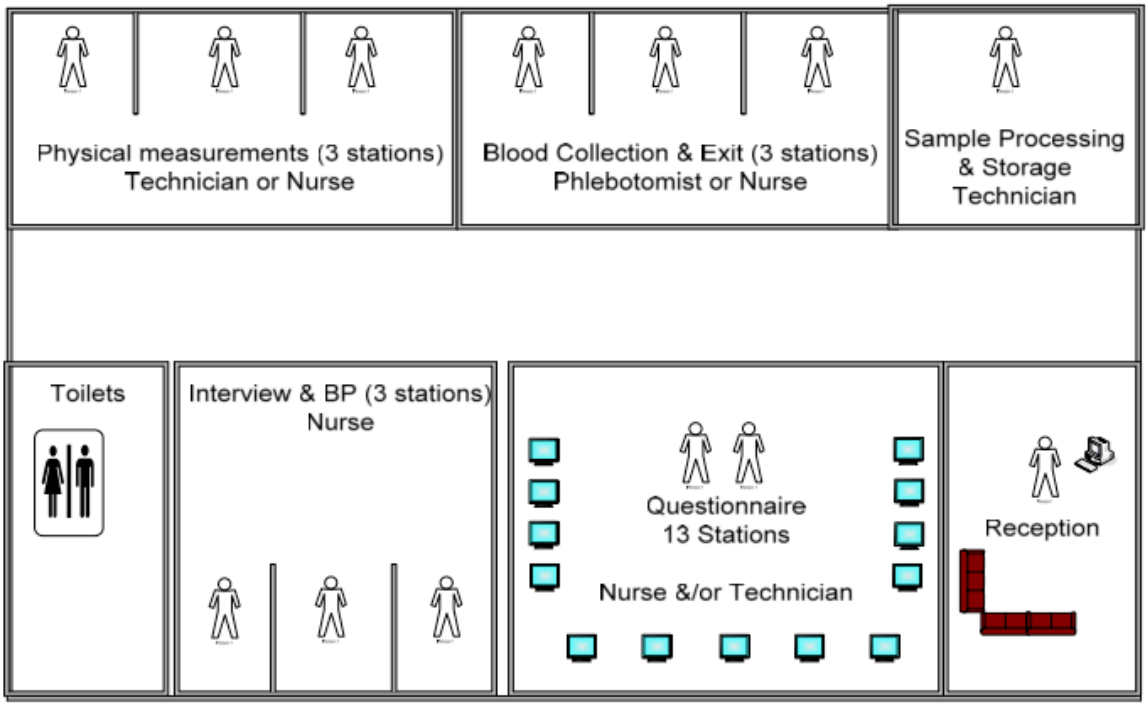



**Figure 3.3:** Overview of assessment centre layout. Sourced from the UK Biobank study protocol.

### 3.2.3. Scope of the self-administered questionnaire

The self-administered questionnaire focused on broad topic areas pertaining to lifestyle and disease risk. These included: sociodemographic and occupation, lifestyle exposures; psychological state; cognitive function, family history of illness, and medical history and general health. Questionnaires previously used in observational studies, clinical trials, and population survey to quantify exposures in above interest areas were used to inform the UKB questions, together with wider consultation with international experts in each area of interest. For most of the topic areas, the questions to select for inclusion in the questionnaire were non-contentious, variable such as the smoking, alcohol, family history, early life exposures, general health and disability were easy to generate and as they have been used and validated in many population studies.

### 3.2.4. Measures relevant to this thesis.

#### 3.2.4.1.Sociodemographic factors

A variety of socioeconomic factors known to be associated with health and wellbeing were assessed using best practice tools. Potential factors that inform socioeconomic class, deprivation, and education formed basis of the social economic measures. This included questions on housing tenure, car ownership, household income and structure, current occupation, ethnicity and country of birth, qualifications and school leaving age. The questions were mostly sourced and adapted from the general population surveys where they have been assessed and validated on large and diverse populations.

#### 3.2.4.2.Smoking and alcohol

Smoking and alcohol measures were included as they are known risk factors for lung and other cancers, cardiovascular diseases, chronic obstructive pulmonary disease, and other number of respiratory conditions. Questions on smoking behaviour were adapted from several longitudinal epidemiological studies and surveys, as well as consultations with experts in the field. Due to the overwhelming evidence about the role of smoking in driving cancers and cardiovascular disease, questions on smoking were comprehensive and covered both duration and frequency of smoking. Alcohol was quantified by means frequency of intake, and included beverage specificity due to some evidence that suggest it may improve under-reporting and it may be an important factor for certain outcomes.

#### 3.2.4.3.Other environment factors

Occupation was collected by trained interviewers with the Standard Occupational Classification 2000 built into the CAPI system (238). This was to allow precise and discriminatory occupational categorisation, and the ability to explore the relevance of this factor as a socioeconomic and environmental determinant disease.

Additionally, many environmental exposures considered to be important predictors of common diseases (respiratory and musculoskeletal conditions and which provided reasonable response distribution were collected using the questionnaire. These included current address, residence at birth, occupation and other workplace factors, residential air pollution and noise pollution. Current address was particularly important as it allowed researchers to explore multiple potential environment risk factors by linkage with UK environmental databases.

#### 3.2.4.4.Physical activity

Questions on physical activity were adapted following a pilot and a validated survey instrument. They were principally intended to allow participants to be ranked according to their levels of physical activity (vigorous, moderate, and walking). Questions on sedentary activities were also included as measures of physical inactivity. Repeat assessment of physical activity were recorded at follow up visit to allow for variation and intensive assessment over time. Derived variable generated from self-reported factors from the International Physical Activity Questionnaire short form (239). This included different physical activity type and duration (such as walking, moderate and vigorous physical activity, strenuous sports…etc). These factors were converted into a single measure of total physical activity in metabolic equivalent of task (MET) – hours per week weighted by intensity (walking, moderate or vigorous)

### 3.2.5. Sources of health-related outcome

Permission to access health outcomes data, past and future were obtained at enrolment from all participants at baseline. These health records were used to supplement information recorded at enrolment about previous medical history, family history, investigations (e.g.: radiology and blood tests), and exposures (medication, occupational health). Most importantly, access to such records was needed to provide follow-up information related to cause-specific mortality and other health events (out-patient and in-patient hospital activity, cancer and other registries and prescribing information). The most reliable single identifier is the NHS number in England and Wales and the Community Health Index

(CHI) number in Scotland. These numbers were obtained for all potential participants prior to their invitation to attend the assessment centre. Other identifiers (such as name, date of birth, address, general practice) were also obtained prior to invitation, and checked during enrolment, to allow linkage to other types of health-related information. These identifiers helped ensure that participants are not lost to follow-up, which may continue for many decades. A variety of different sources and systems were used to ascertain death, disease occurrence and other health-related information among participants during follow-up. Linkage of participants within some of these systems were initiated during the recruitment phase, but linkage to other systems were done later with evolution of the NHS digital systems.

#### 3.2.5.1.Hospital records

Information about UKB health events and procedures experienced by participants when they attend hospital were retrieved for three participating UK countries (England, Wale, and Scotland). For England and Wales, this information is held by the Department of Health's Hospital Episode Statistics (HES). HES is the national statistical data warehouse for England of the care provided by NHS hospitals and for NHS hospital patients treated elsewhere. It is the data source for a wide range of healthcare analysis for the NHS, Government, and many other organisations and individuals. Data held in HES are derived from the NHS-wide Clearing Service that provides the mechanism by which HES data are transferred from individual hospital trusts clinical systems. Each, there are approximately 12 million episodes of care in the database representing all NHS funded admissions for patient care, and private care within the NHS hospitals in England. HES includes information about patient identifier, inpatient, and outpatient episodes, maternity records and psychiatric census, administrative details and the organisation providing the treatment and clinical information relating to diagnoses (ICD10 codes) and procedures (OPCS4 codes).

Historical data is also included which allowed UKB to gain insight each individual past medical history. For Scotland, the Scottish Morbidity Record

(SMR) has been collecting data on all admissions to all Scottish NHS hospitals since 1980, and these data are routinely collated by the information and Statistics Division (ISD) of the Common Services Agency. With the permission of the NHS Privacy Advisory Committee, the UKB was able to extract hospital admissions data for Scotland (and the same structures allowed the retrieval of primary care records, prescribing information, and the maternity, cancer, and death data). Access to identifiable clinical information was obtained via the agreement of the Security and Confidentiality Advisory Group, to allow access to raw codes in specific circumstances. In Scotland, access to the SMR records was provided with the agreement of the NHS Privacy Advisory Committee. Provision of hospital records was acceptable as all participants gave written and signed consent at enrolment for extraction of their individual hospital records and other health related information.

#### 3.2.5.2.Primary care records

Primary care records included a broad range of clinical care information that ranged from organisational records, to detailed care information and pathway. Currently, there is no single national system for collecting or sharing primary care data. UKB consulted with necessary data suppliers and intermediaries to obtain primary care data on all UKB study participants. For Scotland, individual CHI numbers were used to link a range of health-related information, including primary care, clinical and prescribing datasets, which dates to 1984. For England and Wales, Connecting for Health programme, a programme part of the NHS care and Secondary Uses Service. The data also included records from healthcare professionals such as general practitioners, pharmacists, dentists, and opticians. To date, 45% of the UKB cohort (n=231,000) have had their primary care data linked and are available for use and analysis.

The primary care computer systems have different coding classifications for clinical events and prescription. As part of the limited data curation that has been applied, multiple clinical coding is provided per record where these are available. The primary care data were coded using Read codes used in primary care since 1985. There are two versions: version 2 (Read v2) and version 3

(CTV3 and Read v3). Both provide a standard vocabulary for clinicians to record patient findings and procedures. For the purpose of research, the UKB provided key variables important for epidemiological research. These variables included clinical events (diagnoses, history, symptoms, lab results and procedures), prescriptions data and a range of administrative codes.

#### 3.2.5.3.Self-reported health information

Additional information on health events and medication being taken by each participant was captured at baseline and during the follow up. This information served to supplement the data from the hospital and primary care data sources. This method of collecting outcomes data provided an opportunity to collect recent or current conditions (including those that might be under-reported in other data sources) and the medications taken by participant during the follow up period. This data allowed for cross-reference with other information extracted the hospital and primary care records, minimising missing outcomes and helping validate them.

### 3.2.6. Derived outcomes used in this thesis.

#### 3.2.6.1.1. The mapping of the first occurrences data fields in UKB

The mapping of first occurrences data fields involved integrating the routine health data from England, Wales, and Scotland with UKB self-report outcome data. As mentioned earlier, the linked routine data consisted of hospital admission records, coded primary care data, and the death registry. This data represents real-world administrative data whose purpose is to deliver care but not to facilitate research. As such UKB created algorithms (combinations of clinical codes with rules for case inclusion/exclusion, where appropriate for specific outcomes) and this was mapped to the 3-digit coding from the International Classification of Disease (ICD-10) **Figure 3.4**. The algorithms have gone through expert peer review and consensus processes, with information on the positive predictive value and other validation being provided

where possible. In general, there were two main classification systems of clinical coding, ICD and Read codes.

- Hospital inpatient records (ICD-10 and ICD-9)
- Primary care records (Read v2 and Read CTV3)
- Death records (ICD-10)

Hospital inpatient data does not record the date of diagnosis but rather provides information on the date hospital episode started and ended including the date of hospital admissions. The date of first occurrences was defined as the date of first episode for a particular outcome of interest. If this date was missing, either the date of admission, or the episode end date or discharge date were used instead. In the case of primary care data, the date of event was recorded except in the case where this date was improbable like if it corresponded to the participants date of birth. In such cases, events were tagged as being unknown or missing. The dates derived from self-reported information on diagnosis were recorded during a verbal interview at baseline. Participants were asked when they had first been diagnosed (by a doctor) with a condition they self-reported. Participants could provide a year, or their age at diagnosis. These were converted to interpolated year. For example, if the participant gave a year of diagnosis of 1990, then then date of diagnosis was assigned to the middle of the year (06/1990). If they gave the age at diagnosis (30 years old), then the date was set at the year at which they were 30 years old plus 6 months. If the date corresponding to a self-reported medical condition was unknown or had "prefer not to say" response, these participants dates were ignored. If the same 'first occurrence' date was recorded in more than one source field, the source featured first on this ordered list (1.Primary care, 2.Hospital admissions, 3. Self-report, 4.Death) would be used[1].

It is worth noting as with all administrative data, these data sources are subject to a several potential biases related to linkage errors or uncertainty in the linkage and missingness of data (240). The first occurrence date was compared

[1] https://biobank.ndph.ox.ac.uk/showcase/refer.cgi?id=593

with the baseline attendance date to get indication if the condition was prevalent or incident. Dates that were assigned special values (02/02/1902 or 03/03/1903) were also excluded

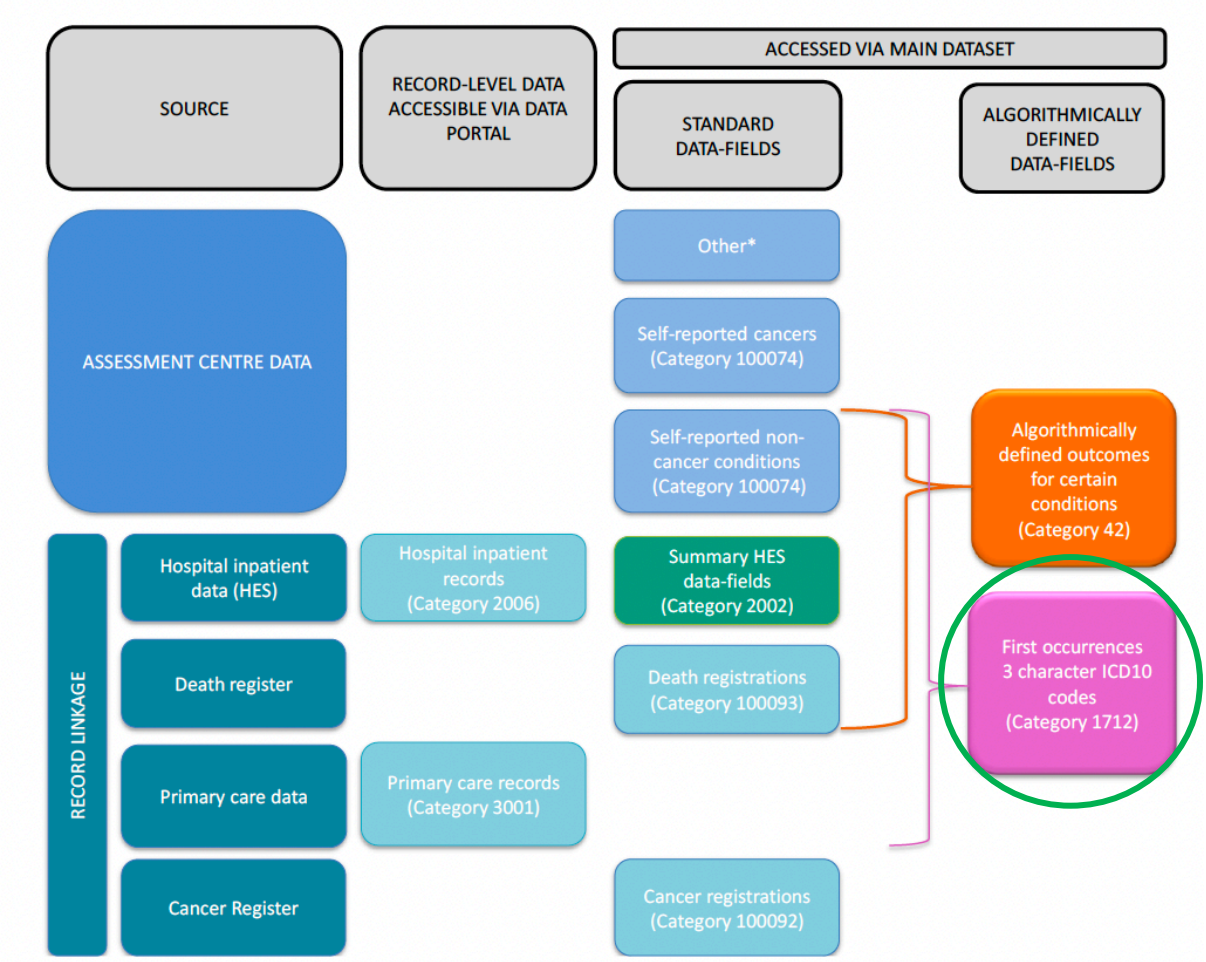



**Figure 3.4:** Sources and method used to derive the first occurrences data field in UKB.

### 3.2.7. Selecting and defining autoimmune diseases

As highlighted above, recorded health outcomes captured either through the hospital or primary care and self-reported questionnaire were mapped centrally by the UKB analyst to the International Classification of Disease (ICD10 and ICD9) chapters. Participants who were diagnosed or who reported to have systemic vasculitis (ICDM30-31), rheumatoid arthritis (ICDM05-06), systemic lupus erythematosus (ICDM32), ankylosing spondylitis (ICDM45) and psoriatic arthritis (ICDM07) were included in this study (Appendix 3). Those assigned more than one of these diagnoses simultaneously at their last date of diagnosis were excluded (n=86)

## 3.3. Description and composition of UK Ireland Vasculitis Registry (UKIVAS)

### 3.3.1. Background

The United Kingdom and Ireland Vasculitis registry is a large database of vasculitis patients from across the UK and Ireland. It is a collaboration between patients, clinicians, and scientists, with a future vision of collecting longitudinal data that will allow for a detailed picture of the clinical course of vasculitis, and providing an evidence base for improved diagnosis and treatment of patients with vasculitis. UKIVAS aims to facilitate and ensure that those working in the field of vasculitis (clinicians, patients, commissioners) have robust and meaningful information that will help drive informed decision making for vasculitis services and patients, taking into the heterogeneity in outcomes across the different centres and regions involved as well as mode of treatment. Importantly, the registry aims to provide researchers working in the field with rich resources that will facilitate large scale studies by enhancing the identification of potential candidates and participants for clinical trials that will improve the lives of patients.

### 3.3.2. Cohort recruitment

Patients who are under regular care at various specialist clinics across the UK and Ireland were recruited since 2009. Over 7,500 patients have consented to be enrolled to the registry and will be followed over time. Included in the registry are information on patient demographics, clinical data related to diagnosis such as signs of organ involvement, ANCA status and overall treatment regimen. The diagnosis of vasculitis was determined by local physician specialist's and using the Chapel Hill Consensus Conference nomenclature of vasculitides criteria.

### 3.3.3. Selecting and defining vasculitis cohort

For this thesis, patients diagnosed with ANCA-Associated vasculitis (GPA, MPA an EGPA) and Giant cell arteritis and who had their postcode data available were included for analysis.

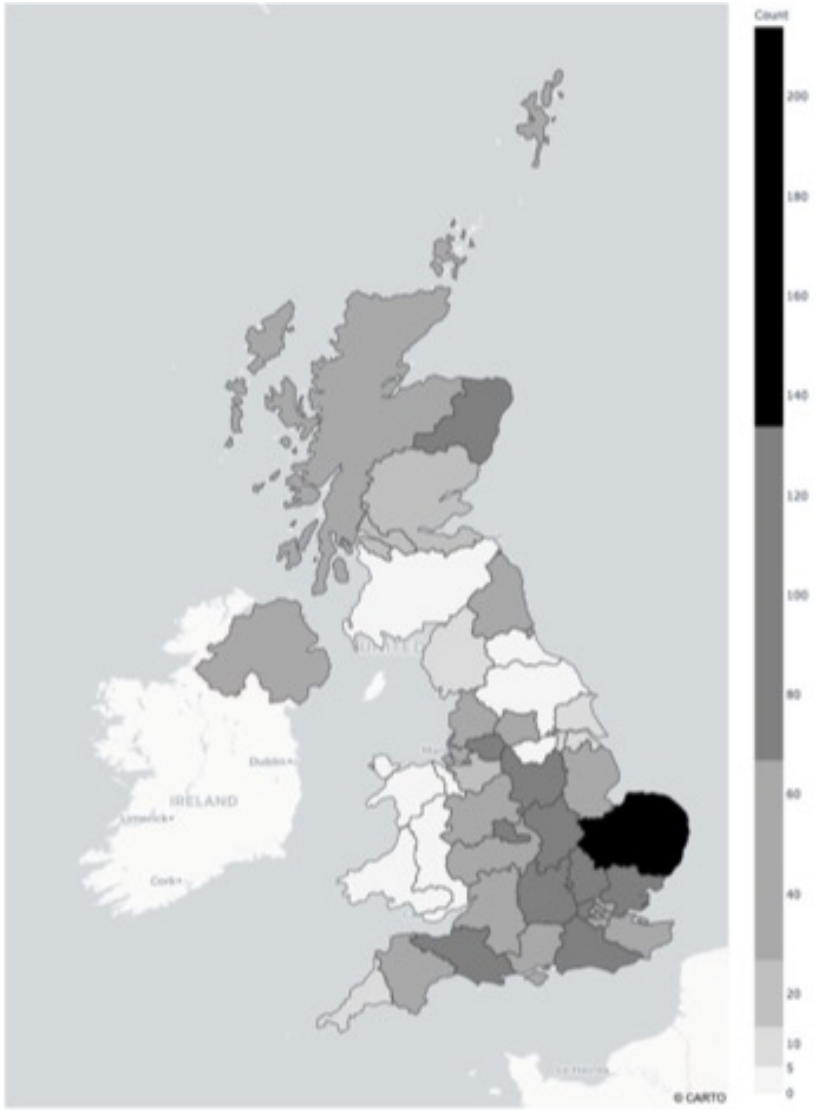



**Figure 3.5**: Distribution of UKIVAS patient location by region
Source: Scott and Havyarimana et al.,2022.

## 3.4. Description of the Scottish vasculitis cohort

### 3.4.1. Background

The data that will be used to address the second and third objective of this thesis and that will be used in chapter 6 and 7 of the thesis were sourced from the Information Services Division (ISD), a division of the National Services Scotland (NSS), part of the National Health Services (NHS) and Public Health Scotland (PHS). The ISD hosts the National Data Catalogue (NDC) which serves as the single definitive resource of information on Scottish health and social care. Its vision is to improve the provision of comprehensive national data through a single point of access, eliminating duplication of information and enhancing support for social and healthcare research. The ISD produces a wide range of datasets that contributes to the understanding of health of Scotland's population. It hosts mainly information on population composition, material deprivation, death, cause of Suicide as well as the hospital admission, and primary care records.

This resource contains health information for over 5 million people in Scotland. In some cases, it covers data on individual's whole life (from before birth using the mother's antenatal record, through to the individual's death). The data are a great tool for looking at patient pathways and follow-up, including hospital admissions and survival and is available for linkage and research. Data linkage of multiple data resources hosted by NSS was achieved by using the community health index (CHI). The CHI is a computer-based population index whose main function is to support primary services. It contains details of all Scottish residents registered with a General Practitioner and is intended to be the Scottish equivalent of the new NHS number in England and Wales and should serve as the unique patient identifier throughout NHS Scotland. For this thesis we used mainly data from the Scottish Morbidity record (SMR01). The SMR01 holds data on inpatient and day case hospital discharges from non-obstetric and non-psychiatric hospitals in Scotland. Specifically, these are hospital episode data collected at admission and discharge, and with the purpose of providing clinical service, including monitoring wait times and

duration of care. For instance, a single admission may contain multiple episodes if a patient is transferred to another consultant and speciality. The main condition of treatment and up to five other conditions may be recorded for each episode and is coded using the International Classification of Disease codes (ICD-10). It is possible for different episodes within one admission to have different main conditions

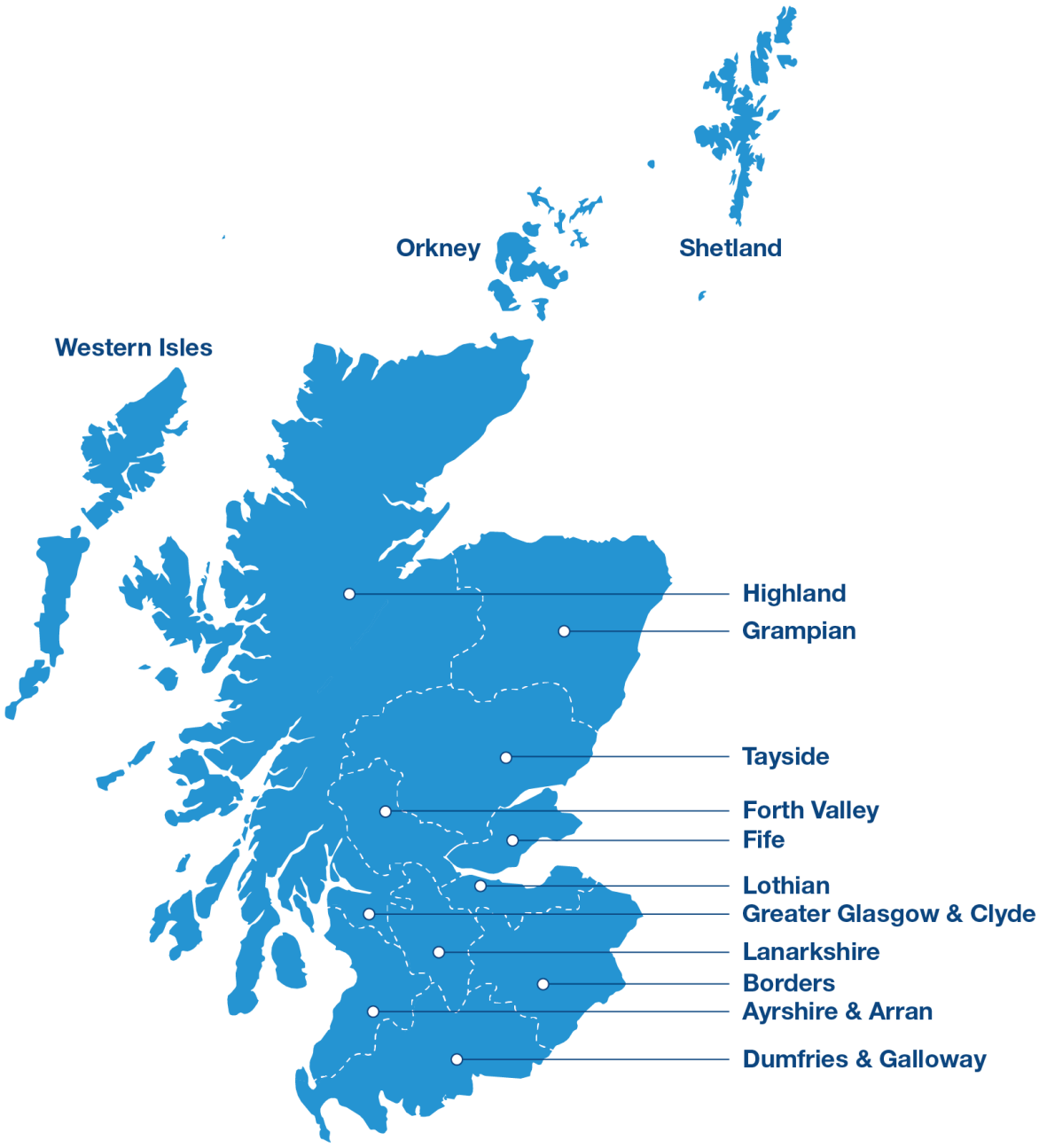



**Figure 3.6**: Service map showing the 14 health boards of care and treatment in Scotland

### 3.4.2. Study population

We focused on AAV and GCA patients; due to the very small number of cases of other systemic vasculitides which are potentially identifiable and therefore would not be disclosed by ISD. Estimated prevalence of systemic vasculitis ranges from 0.1 per 100 000 to 2.6 per 100 000 population for ANCA-associated vasculitis (AAV) and 25 per 100,000 (Giant Cell Arteritis) and to date more than 1,500 AAV and GCA patients have been identified by the ISD team. All patients with AAV and GCA in Scotland were identified from the national administrative databases (SMR01) with a control group comprising of general population controls sourced from the CHI register. Cases and controls were matched by age and sex and included all adults (aged ≥16 years old on index date). ICD-10 codes, used in Scotland since April 1996, provide greater certainty and specificity than ICD-9 codes. AAV and GCA were defined by at least one of the following ICD 10 codes present during the study period (1996-2021), Granulomatosis with Polyangiitis (GPA - M31.3), Microscopic Polyangiitis (MPA - M31.7),Eosinophilic granulomatosis with polyangiitis [Churg-Strauss]/EGPA - M30.1), Giant Cell Arteritis (GCA - M31.5 and M31.6), Polyarteritis nodosa (M30.0),Other conditions related to polyarteritis nodosa (M30.8)

### 3.4.3. Data linkage

Two types of data linkage were carried, one, for electronic administrative health record and the second, for the environmental data linkage. The linking of administrative health records held by the ISD was facilitated by the electronic Data Research and Innovation Service (eDRIS). In line with our research goals, the eDRIS provided the linkage of the following datasets consisting of important measures and confounding measures which need to be considered when adjusting for the impact of environmental factors on systemic vasculitis:

(a) Scottish Morbidity Record (SMR) – records of all hospital inpatient and day case admissions, (SMR01), cancer registry (SMR06) and mortality.
(b) CHI register– enabled matching of the control group and provide confirmation of GP registration, and was used to determine loss to follow up

(c) NRS Death Register (held in eDRIS) – mortality register
(d) Scottish Index of Multiple Deprivation – Deprivation Measures
(e) Scottish Government Urban Rural Classification – Population Density

The de-identified linked data were analysed in the National Safe haven and is retained by eDRIS for the duration of study period.
To enable assessment of linkage and data extraction quality, validation metrics including duplicate extraction of small number of variables across datasets (e.g.: gender) and expected pairings (disease plus prescription pairing) were used as positive and negative controls for the linkage process.

The environmental linkage was achieved transforming the patient's postcode into grid coordinates (easting and northing) and linking with annual pollution measures, which were modelled estimates using the Pollution Climate Mapping models) developed by UK department for Environment, Food and Rural Affairs (DEFRA) (https://uk-air.defra.gov.uk/data/pcm-data). In brief air pollution measure was estimated on a 1kmx1km square grid which has an associated central point (centroid). The estimated pollution concentration provided were calculate at an intermediate zone level around the centroid. A mean estimate for that zone is then given and patient within that zone were assigned a measure of exposure based on the year of diagnosis. For zones where there were no grid square centroids, the pollution estimates from the nearest grid square were used. Similar approach was used for the weather data sourced from the UK Met office (https://www.metoffice.gov.uk/research/climate/maps-and-data/data/haduk-grid/datasets).

# Chapter 4 – Environmental data linkage for clinical benefit

## 4.1. Overview

The following chapter will discuss methodologies and approaches used to link environmental data with routine and clinical health data. It will provide a brief history of record linkage of routine data and the use case for research. It will lay out the benefits and limitation of such data for research as well use for investigating environmental aetiologies associated with diseases and health. Specific methodology related to data linkage of environmental data with routine health are also discussed. This is followed by a pilot study evaluating the feasibility of linking clinical registry data with biobank to investigate the impact of environmental exposure on vasculitis and ends with discussion of consideration points of the pilot study and how they inform later chapters addressing thesis objective.

## 4.2. History of record linkage in the context of health research

William Farr (1874) was the first to recognise the benefits of bringing together separately held records pertaining to important health events (241). He recognised that these data could be used for research and to conduct cross-sectional and longitudinal studies that would improve our understanding of the risk factors associated with morbidities and mortalities across England and Wales (242). Such recognition could not be achieved at the time of his publication. There was limited ability to conduct any type of record linkage. The idea of merging two large record data at a population level did not gain much attention until Stocks (1944) and Dunn (1946) reiterated its advantages (242,243). Dunn proposed the idea that, each person creates a “Book of Life”, starting with birth, ending in death, and composed of records of important health and social events(243). He developed the concept that it was important to collate personal records into an individual file and would later name this process

record linkage. He predicted that statistical analysis of linked records would be useful to health and welfare agencies in evaluating their service programmes. He also suggested that linking a person's individual record would help establish the accuracy of the source and provide a historic trajectory of a person's life. With this recognition, many attempted record linkage, first through manual collation of vital events from birth certificates into single files. Canada was the first to attempt this in 1947(244). Soon after, Newcombe and Kennedy (1959) added important methodological contributions that would allow the linking of stillbirth and birth certificates with marriage certificates leading to one of the first genealogical mapping of the British Columbia population in Canada(245). At the same time, Axford and James (1959) took advantage of the technological revolution and rise in healthcare records and showed that it was possible to programme a computer to link family groups and their corresponding data; speeding the record linkage process, despite the discrepancies that may exist in the data (246). Part of this recognition came after the inception of the UK National Health Services (NHS) in 1948(247). During this time, it was recognised that more health information was needed to evaluate the prognostic, survival and reoccurrence of health outcomes affecting the public, while improving the overall management of the NHS(242). Since then, the recent advances in information technology and data platforms across the globe have provided tremendous promise for the use of administrative health data including social and environmental repositories to drive public health research. Example of this can be seen with the substantial rise in the number of studies using these data to investigate trends and risk factors associated with chronic diseases(248–250). Furthermore, the recent discovery and detailed classification of rare diseases like ANCA vasculitis (AAV) and Giant cell arteritis (GCA) and their growing presence and prevalence in these administrative health data is still under-appreciated and researched in the UK and worldwide. It is this realisation that drives many of the questions addressed in this thesis. This thesis and the result chapters will attempt to answer key questions aimed at bringing new knowledge and evidence about possible new risk factors of vasculitis. This method chapter will highlight the role of modern linkage methods that can help us address these questions and improve our understanding of environmental triggers of vasculitis in UK.

## 4.3. Benefits and limitations of using routine data for research

### 4.3.1. Benefits

There are many advantages to using routine data for research and to inform new scientific evidence that can shape public health policies and improve our understanding of disease aetiologies. Analyses of these data can answer questions requiring statistical power or large sample sizes, improve research coverage of hard-to-reach populations and provide high level validity that can inform policy making (251). Routine data allow for the ability to undertake cost-efficient research that is affordable and easy to implement when compared with consent-based clinical studies involving the recruitment of study participants. Through record linkage, these data can provide substantial return on investment which hasn't always been appreciated. Loss to follow up is a problem in clinical studies as it is defined by time and frequency of visit by study participants. Through linkage, attrition bias can be reduced as those lost to follow up can often be different to those remaining in care. There is also added value through improvement of data quality. For example, multiple errors are picked up and corrected during data linkage, this including any other technical challenges. This process leads to greater accuracy of records and methodology used for record keeping (252). Contrasted to consent based approaches, linkage studies can be better for conserving patient privacy regardless of whether they would have given consent to the use of information (253). For example, approval panels responsible for upholding the legal basis of disclosing routine data, are often looking whether the proposed research is in the public interest and benefit, taking to account whether the benefits outweigh any potential risk for using these data (254). In the United Kingdom, approval panels assess applications based on principles from the Data protection Act, 2018 of 'fair and lawful basic for processing data' (Information Commissioner's Office, 2018) (255).Lastly, by bringing different data sources together, record linkage provides a focal point for envisioning future possibilities, better cooperation, and rigorous debate about the uses of data and the results in subsequent research.

A good example of this is UK Biobank. UK Biobank brought together routine data from national cohorts (England, Wales, and Scotland) with separate data providers and coding systems to provide a rich resource that have allowed for detailed investigation into genetic and environmental risk factors associated with heath and disease (256). Such linkage can strengthen collaboration between multidisciplinary teams (clinicians, administrators, researchers, patient groups and the public), and encourages the necessary teamwork needed to tackle health and social problems. Further, publications derived from linkage projects provide important real-world evidence that can shape better care, prevention strategies and treatment programs including policy formation and service delivery reforms (251). Several examples of this can be seen in references of policies addressing health inequalities in the UK population, in particular policies on drug and alcohol use as well as clean air strategies for air pollution (257–259).

### 4.3.2. Limitations

Limitations of administrative data are well documented in the literature, particularly those related to data quality and missing data (39,251,260). For example, missingness of data can occur if a person fails to interact with the service, especially in hard-to-reach population and individuals with poor health-seeking behaviour. Many administrative data contain inconsistency or incomplete data that vary in structure, format, and content, especially with repeat observations (261). Data pre-processing therefore is time-consuming but is essential to achieve quality of linkage. Research applications and data access can also take long time where the application processes are not streamline. An example is in the case where approvals from different bodies are required which could lead to same information being share with different panel. Such instances can cause delays in application process and may hamper project schedules (262). Importantly, linkage errors can affect the quality of the data especially where records from the same individual fail to link as it happens with typographical errors, changes in person's characteristics (e.g.: changes to a person's name after marriage or address) or in cases where there are incomplete characters or values in the linkage key (263). Such errors can

threaten the reliability of results though it may be minimised if effective communication between data providers, linkers and analysts is in place.

## 4.4. Use of routine data for environmental health research

The use of routine health care data for research has shown to be important for providing evidence on aetiologies of many diseases, inform better care and management of patients as well as shape environmental health policies (264). The shaping of environmental health policies has been possible because of the advancements in environmental monitoring tools and methodologies which have been established to monitor the impact of the natural and built environment on people's health and wider environment (265). A by-product of availing such datasets (particularly environmental monitoring data and routine health data) allows many researchers to study the long-term impact of outdoor air environment on population health and disease. This is especially true for diseases that are common in the general population. Similar approaches are being applied to rare diseases where major gaps still exist about their risk factors (197). Furthermore, the growth of these datasets in terms sizes and coverage have meant that more studies are being conducted and allow for investigation into spatial and temporal clusters of diseases and the relationship with environmental exposures. For instance, the past ten years have seen a rise in number of studies that show air pollution to be an important risk factor of cardio-respiratory diseases and that its effect may vary based on geography and deprivation status (181,266). Additionally, some health events have uneven distribution across time, with some events (e.g.: rates of respiratory disease admissions) being seen to increase at regular intervals (e.g.: higher admissions in winter) while others at irregular intervals with environmental exposures being a leading predictor of these trend (267,268). Such investigations are possible because of the many linkage tools and methods available to integrate environmental monitoring data with routine health data based on location and time of event. These methods are summarised below together with necessary consideration with regards to how one method may be chosen over the other. The decision to use a specific method will depend on

how environmental monitoring data are recorded and presented across the population. For example, data from environmental monitoring stations are not equally distributed across urban and rural geographies (269). The representativeness of these exposure estimates may not accurately reflect real exposure at population level. This is particularly common if the data is provided at shorter time scales (daily and monthly). In some cases, environmental data is provided as modelled estimates derived from pollution climate models (270). These types of exposure metrics are calculated at annual scales and have been used in Chapter 5 and 6. All these exposure estimates, whether modelled or from monitoring stations, can be linked to clinical data based on an individual's postal code . In some instances, patient postcodes may not be available and only location information about the primary and tertiary care provide is available. In such instances, appropriate linkage method should be used and interpreted with care. Further details about exposure assessment approaches are discussed below and the corresponding linkage consideration which will further be detailed in the result chapters.

## 4.5. Use of proximal vs indirect exposure assessment in aetiology research

As highlighted above, the decision on how to assign environmental exposure for individuals in a research study is dependent on the type of environmental measures that can be accessed and the capacity to undertake specialised linkages. Ideally, to investigate the effect of environmental exposures on health outcomes requires the use of quantitative exposure measures that are accurate, precise, and clinically relevant, preferably covering a range of exposure doses for a given population. In reality, these ideals are often not met, and exposure measures do not reflect 'true exposure', or dose and in many cases, vary by methodology and tools used collect and estimate them. Examples of environmental measures used in research can be grouped into two categories: proximal based exposure estimation and indirect exposure estimation. Indirect exposure estimation was introduced in chapter 1 and is further detailed in this chapter and will be used in the result chapters.

Proximal estimation of environmental exposure involves the use of distance-base metrics that are calculated based on individuals proximity to areas with high air pollution, such as the distance from major roads with increased traffic intensity or industrial areas (271–273). Other major examples can be seen in studies where access to individual postcode data is restricted or not available. In such cases proximal estimates are derived from ground monitoring station data nearest to the patient's hospital or primary care physician (Figure 4.1). Patients are often grouped based on spatial hierarchies, for example, the lower or medium super output area in the UK, or county in the US, or prefecture in Japan. Such grouping allows for comparison of disease clusters, based on health boards while relating them to environmental exposures.

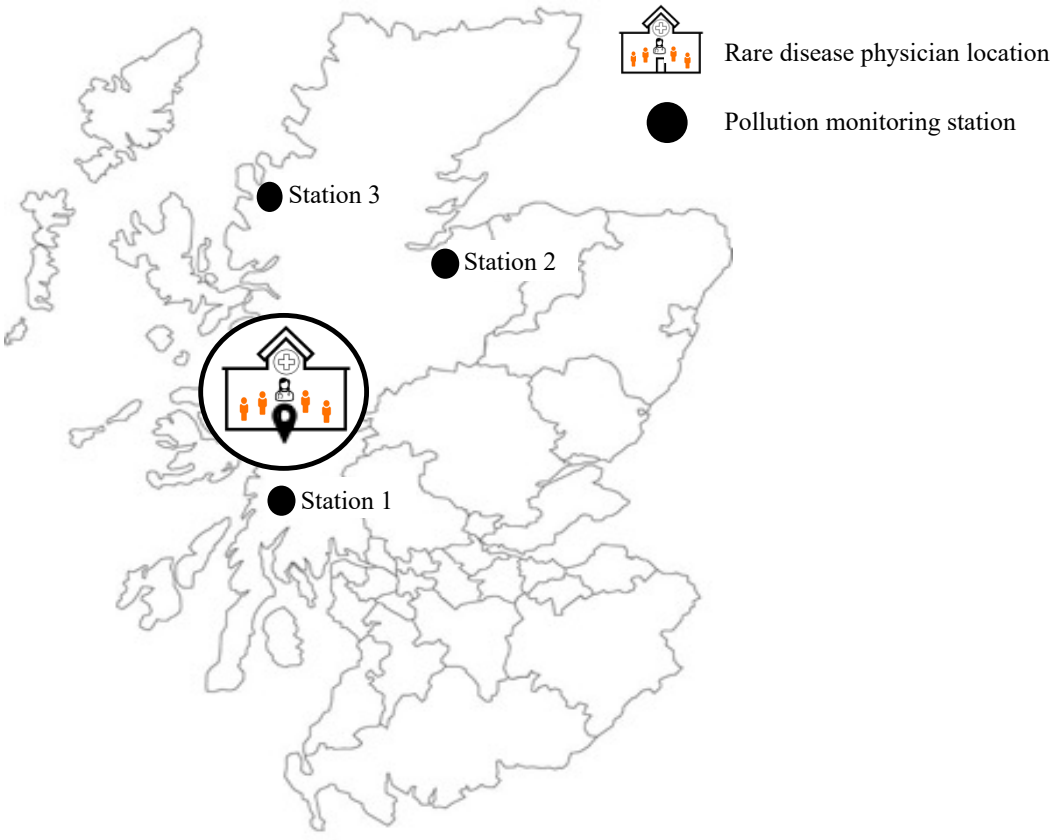



**Figure 4.1:** Schematic example of proximal based approach to exposure assessment. [*map of Scotland*]

Indirect exposure estimates are often achieved in studies where explicit consent or panel approvals based on the legal basis for the use of study

participant's address. In this cases, indirect exposure measures are derived from interpolated exposure data that is modelled at 1x1km spatial grid scale and calculated using land use regression or neural network modelling techniques (Figure 4.2) (270,274,275). The models are calculated using input data from multiple source. These data include ground monitoring station data for air pollution, weather, and atmospheric measure from satellite as well as other measures related to the built environment (e.g.: population density). The provision of such exposure estimates at granular spatial scale (1kmx1km) are important especially for nested case controls studies where differences in background air pollution exposure are needed for when estimating relative risk associated with a particular disease or outcome.

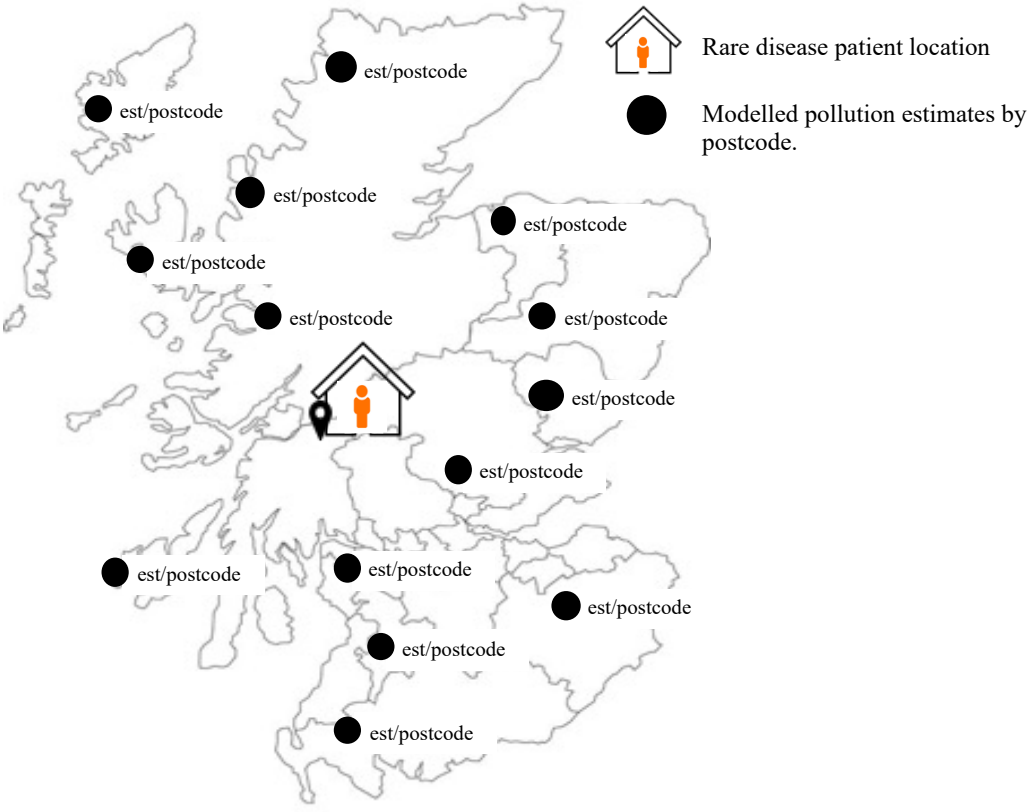



**Figure 4.2:** Schematic example of the indirect exposure estimation assessment. [*map of Scotland*]

## 4.6. Sampling strategy and coverage

Sampling for rare disease that is not prevalent in the general population requires the need to be as precise as possible compared to sampling for a more prevalent disease such cardiovascular disease. In this case, multiple sampling approaches are needed to allow for adequate sample size. The sampling frame should incorporate both probabilistic and non-probabilistic sampling, while aiming to minimise selection bias (276). Convenient sampling (non-probabilistic) is often used when selecting vasculitis cases. This is due the low prevalence of cases and the need to include as many patients as possible. Random sampling (Probabilistic) will often be used to select a control group (Figure 4.3).

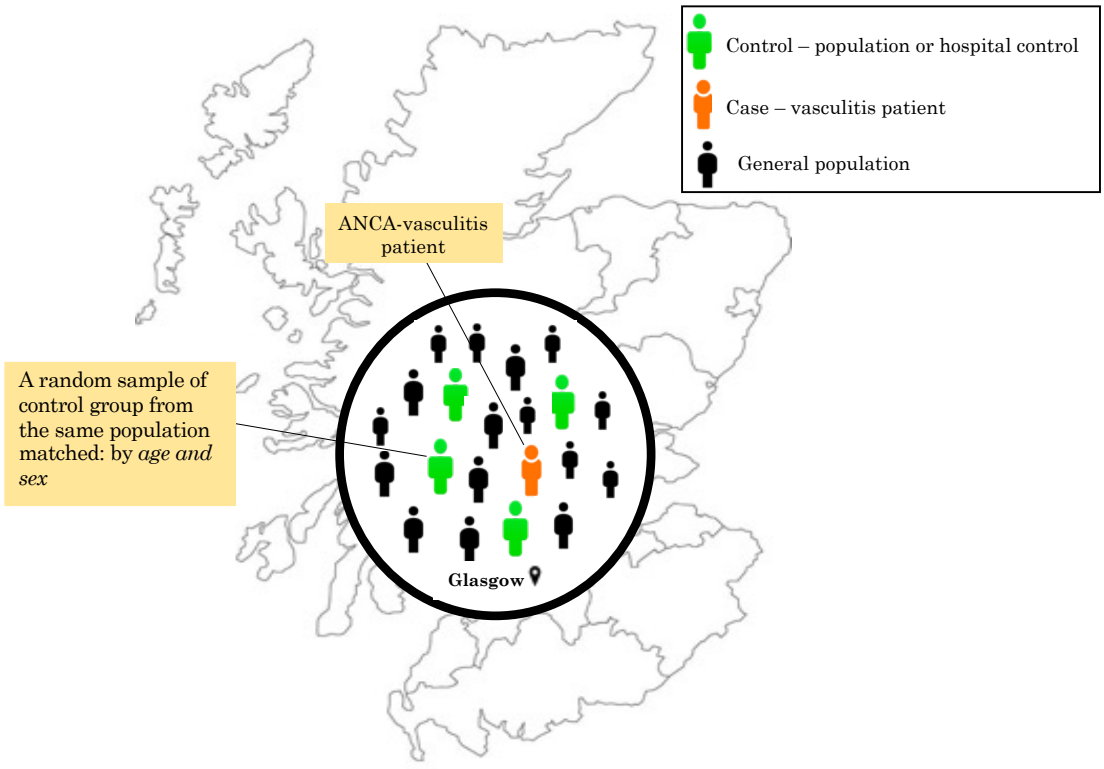



**Figure 4.3:** Schematic example of the random sampling approach used in vasculitis research [*map of Scotland*]

In many of these cases, patients are recruited at a single specialist hospital or clinic by treating physicians. In other cases, patients are retrieved from administrative health datasets, where all patients with a specific diagnostic code

for vasculitis are selected and included for analysis. A control group is retrieved from hospital registry or general population census. The latter sampling approach allow for broad coverage and inclusion of patients from the wider population and covers hard to reach populations such as those remote and rural areas. This is important as these patient are underrepresented in nested case control studies, due to limited specialised care services where recruitment often takes place. Random sampling based on age, sex and geography is also important to allow for generalisability of results for the population at hand, in the case of vasculitis, people who are 50 years or older.

## 4.7. Record linkage methods for environmental health research

### 4.7.1. Background

There are several ways of linking and estimating exposure values based on routinely collected data from environmental agencies like the UK department for environment and rural affairs (DEFRA) and the meteorological office (Met office) (270,277). The rationale for using one method over another, the significance of each exposure value and their effect on the research question at hand are factors that are often not explained by many researchers. When using data from monitoring stations, there are factors that can complicate the computation of exposure values. These include large variation in pollution or temperature readings that are dependent on the ground elevation where instruments are placed (278). The design of the monitoring network and the density of recording stations can all impact the reliability of exposure estimates. For instance, areas that have high density of monitoring stations, like urban areas, provide more reliable pollution or temperature estimates when compared with those from remote areas (279,280). The way the monitoring sites are set up throughout an area can also affect exposure estimates, with reliable estimates being expected from many sites, but this still depends on whether they are evenly distributed or clustered. Importantly, when assessing the impact of an exposure on health outcomes, the distribution of the population within an area can affect the overall risk estimation and inference especially for less

populated areas. Up to now, many researchers have used non-computationally intensive methods to calculate pollution and weather exposure estimates(281). These included using the mean value of all station's recordings within a chosen geographic radius. Another similar approach is to calculate the mean values from the nearest neighbouring stations. These methods have shown popularity over the years. However today, they are slowly being replaced with more advanced and sophisticated methods that focuses on distance weighted averages that take to account census and population density distribution in an area of interest (281–284). Additionally, in air pollution research, there is effort to minimise missingness of data in areas where monitoring station are sparse. This is done by providing researchers with modelled estimates derived from dispersion models or land use regressions. Some researchers still prefer using raw data from monitoring stations and assign exposure values based on polygon map centroids. It is worth noting that there is no clear consensus on which method is best to use. Some methods are computationally intensive while others that are commercially available are expensive and may not be user friendly (283).

### 4.7.2. Linkage options for disease-exposure research

The majority of environmental data provided to researchers are given in a form of gridded map. Several methods have been developed to allow a linkage of this data for assigning of exposure values to individuals in a particular study (281,285,286). A summary of popular methods and their exposure estimation approach are discussed below. These include exposure estimates assigned based on:

i) Average of internal nearest neighbour stations using intersecting polygon
ii) Average of nearest neighbour stations using intersecting proximity polygons
iii) Geographic centroid inverse distance weighted average (using stations that are with 10-50km from centroid)

iv) Population centroid inverse distance weighted average (using stations 10-50km from the centroid)

**i) Average of internal nearest neighbour stations using intersecting polygon**

In this method the mean values from monitoring stations within a chosen polygon, or the nearest neighbour polygon are calculated and assigned to each study participant or patient. The following steps are used, first, the starting point is to identify pollution or weather stations that are within a patient boundary (e.g.: 1km from the participants address or lower super output area). If no stations are available within this boundary, the nearest neighbour stations data are used. The nearest neighbours are found by overlaying and intersecting the proximity polygons (also known as Thiessen or Voronoi polygons) with LSOA. With this approach, each monitoring station serves as the focal point to calculate boundaries of proximity whose area is defined based on the polygon structure. All other monitoring station and points nearest to the focal point serve as a nearest neighbour. Finally, the mean daily or monthly or yearly observation from each of the stations are calculated for each of the boundary and serve as point estimate exposure for a patient or comparator within that boundary.

**ii) Average of nearest neighbour stations using intersecting proximity polygons**

This approach calculates the average of the “nearest neighbours” regardless of their location inside or outside each LSOA boundary. Proximity polygons are used to allocate estimates from the nearest neighbours as described in the first method.

**iii) Geographic centroid inverse distance weighted average (using stations that are within 10-50km from centroid)**

The third option is to use the distance between the geographic centroid of each station by calculate the inverse distance weighted average. This is done by establishing the geographic centroid (known as the mean centre), this serves

as the geographic centre for a boundary of interest. The inverse distances from this centre are used to weight the mean value from each the monitoring stations observations. An arbitrary maximum distance of 10, 25 or 50km from the centroid of each spatial unit is used because it is likely that stations that are beyond this distance will not be like those from an area of interests

**iv) Population centroid inverse distance weighted average (using stations 10-50km away from the centroid**

The last approach is to use the station or estimated measurements based on distance between the LSOA population weighted centroid and the stations for an inverse distance weighted average. The population weighted centroid is calculated by splitting and dividing a LSOA grid into its population census constituents' sub-units. The centroid for each of these subunits is then calculated. The population-weighted centroid is found by weighting the average of the latitude and longitude coordinates of the sub-unit centre, by the populations of those sub-units.

## 4.8. A case in point - Linking environmental exposure data with UKIVAS registry: A pilot study

### 4.8.1. Background

As highlighted earlier chapter 1, one of the main objective of this thesis is to investigate the impact of environmental exposures on vasculitis and to provide new evidence on potential candidate triggers associated with AAV and GCA onset. To achieve this, this thesis is using multiple data sources with patient data of people living with vasculitis. One of the major sources containing a large sample population of patients with vasculitis in the UK is UKIVAS and UK Biobank (Figure 4.4). UKIVAS registry is a clinician led register that collect robust clinical data to facilitate large scale research that can improve patients' lives and improve our understanding of potential risk factors of vasculitis. The registry holds clinical information on the diagnosis and management of patients

with vasculitis. This includes data on variables such as the date of diagnosis, postcode, and ANCA serotype. By using information about the patient address and time of diagnosis, a pilot study was conceived to assess the possibility of merging a vasculitis registry with a population biobank data to facilitate research into environmental triggers of vasculitis. To make the pilot study meaningful, a case-control design was used by merging AAV cases from UKIVAS with unmatched population control group from UK Biobank (Figure 4.5). This was possible because both data sources cover the same UK population in terms of geography and age (e.g.: people mostly over the age of 40 years)

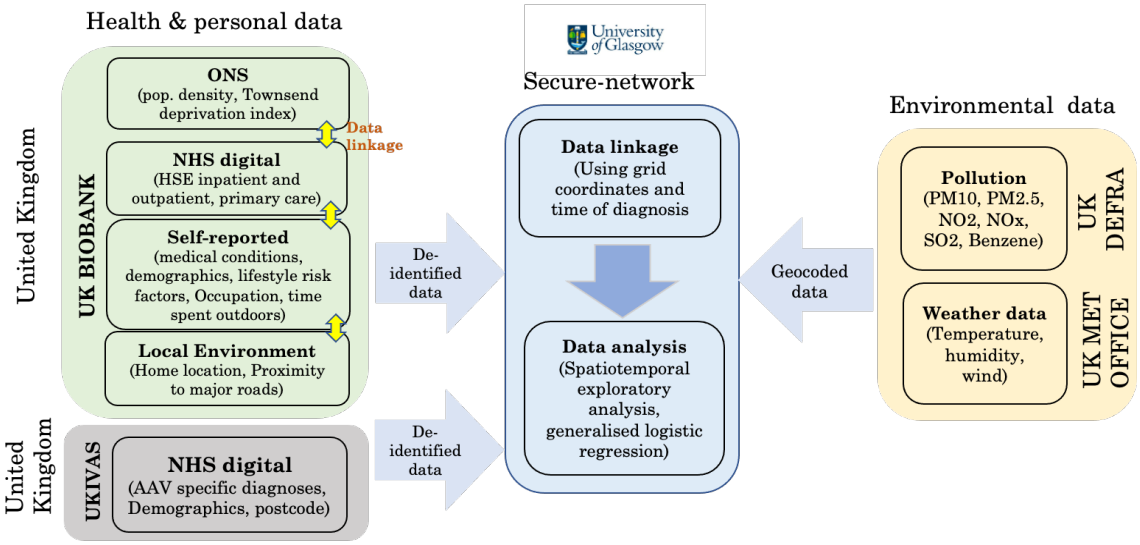



**Figure 4.4:** Overview of the data sources used for the UKIVAS linkage and pilot study

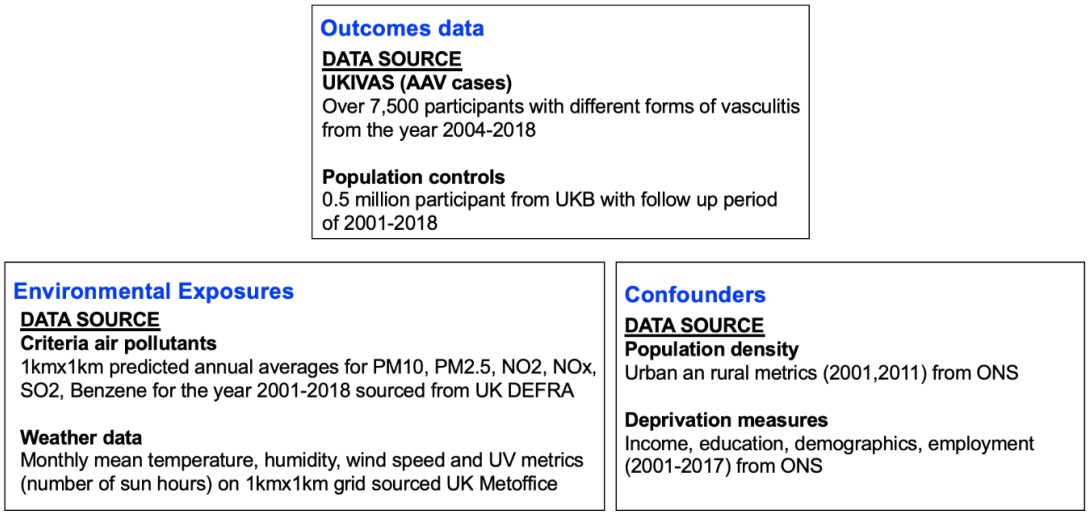



**Figure 4.5:** Overview of the study design used for the UKIVAS linkage and exposure estimation

The merging of the two-health repository can be classified as a form of indirect linkage, which was achieved by merging on common variables (age, sex, ethnicity). As highlighted earlier in chapter 3, UK Biobank enrolled over 0.5 million participants between the age 37 and 70. Each participant consented to have their health, personal and geographic data linked to UKB repository at baseline and follow up. To maximise on participant privacy, postcode data were converted to easting and northing coordinates (at 1km grid accuracy) before being released to researchers. These coordinate data together with information about participants outcomes from the first occurrences data field were used to link air pollution data to each participant's residential address. The selection of UKB control group was based on its representativeness of the population at risk (people over the age of 50 with a non-inflammatory diseases). More details about the selection criteria are provided in the paragraph below. Both the indirect linkage and environmental linkage of UKIVAS and UK Biobank was conducted inside the university of Glasgow secure network (Figure 4.5).

### 4.8.2. Linkage approach and eligibility criteria

An unmatched control group consisting of participants diagnosed with a non-inflammatory rheumatic disease (ICD M15-19) between 2004 and 2018 were used from UKB in this pilot study. This group was selected because it mimics a hospital control group and is also representative of the population at risk in terms of age and sex (people over the age of 50 and female). The environmental linkage was achieved using indirect exposure estimates provided by the UK department for environment, food, and rural affairs (DEFRA) and the UK Meteorological office (Metoffice). Each AAV case and UKB control were assigned a modelled pollution estimates from 2004 to 2018 using the internal nearest estimates from the intersecting polygon (Figure 4.6) (285). A similar approach was used for the linkage of weather variables. This linkage involved the use of ground monitoring station data. As part of the data pre-processing, cases who resided outside the 22 cities of UKB centre[2] and

[2] https://biobank.ndph.ox.ac.uk/ukb/field.cgi?id=54

recruitment boundaries were excluded in the linkage (Figure 4.7). For example, AAV cases from north of Scotland and East of England were excluded. Table 4.1 and Table 4.2 provide a detailed summary of the format of the linked variables, their data sources and the general characteristics of participants who had environmental data linked to their place of residence.

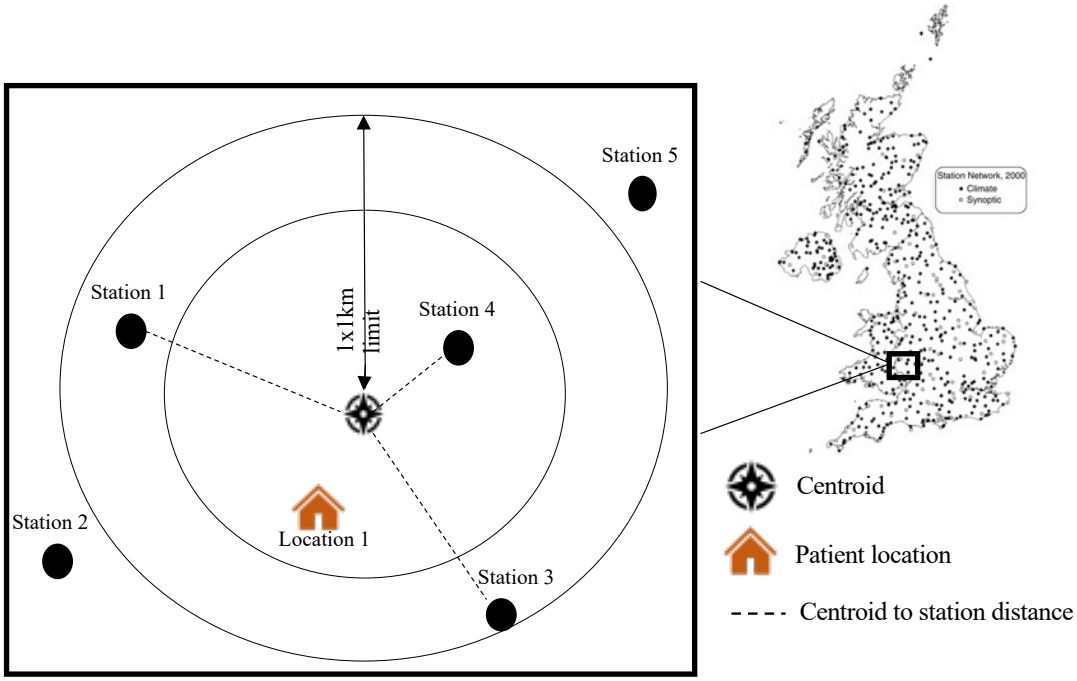



**Figure 4.6:** A schematic overview of the UK weather monitoring network (Perry and Hollis, 2005) and how this data was linked to patient location in UKIVAS.

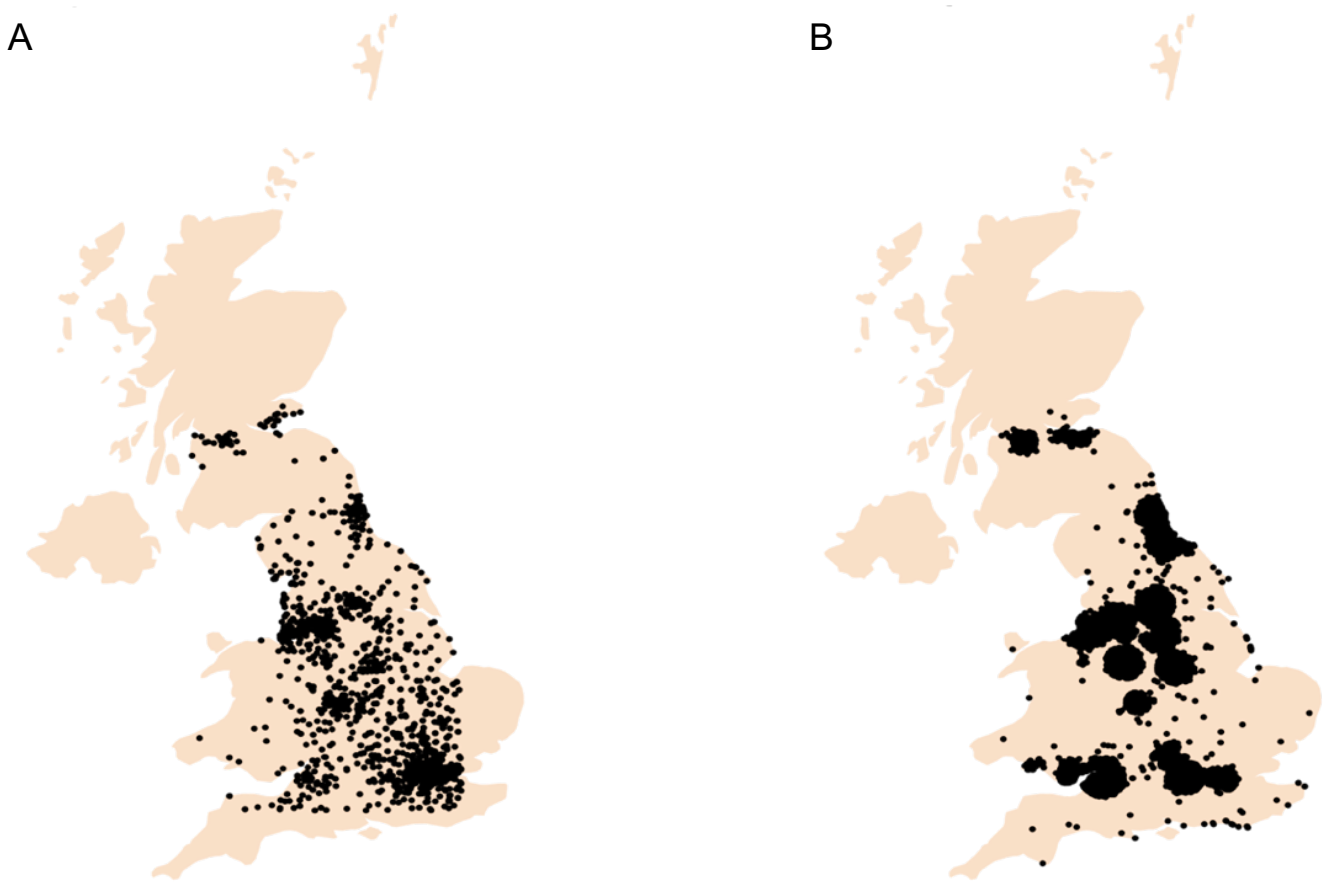
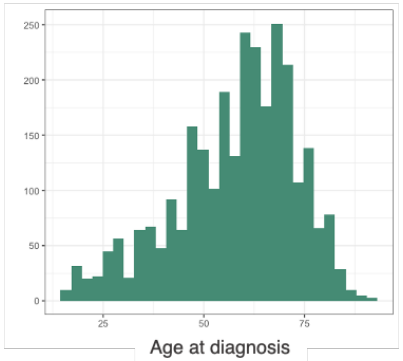



**Figure 4.7:** A map showing the distribution of ANCA-associated vasculitis cases (A) from UKIVAS and unmatched control group from UK Biobank (B). Each dot represents a study participant who were included in the environmental linkage

**Table 4.1:** Descriptive overview of the demographic variables from UKIVAS registry

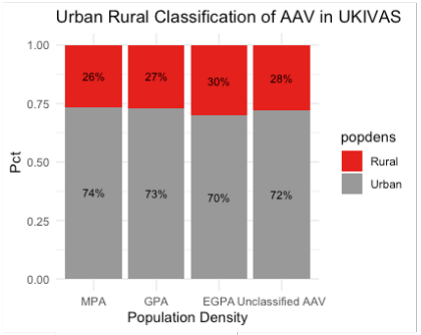
| UKIVAS clinical variables | | | | | | Linked variables | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Demographic | | Confounders | | Weather | | Pollution | |
| | | | | | | | | Month of diagnosis | | 6-year avg. before the date of diagnosis | |
| Row | ID | Date of diagnosis | Diagnosis | Age | Sex | Pop. dens | Deprivation (IMD) | Temperature (Celsius) | Humidity (%) | PM2.5 µg/m3 | SO2 µg/m3 |
| 1 | ID 1 | 12/01/2004 | AAV | 46 | F | Urban | 23.4 | 4.4 °C | 70% | 22.2 µg/m3 | 0.8 µg/m3 |
| 2 | ID 2 | 08/07/2008 | AAV | 55 | F | Rural | 16.1 | 15.6 °C | 30% | 21.1 µg/m3 | 5.2 µg/m3 |
| 3 | ID 3 | 12/03/2015 | AAV | 59 | M | Rural | 9.9 | 7.7 °C | 45% | 15.6 µg/m3 | 8.3 µg/m3 |

Age at index [3]




UKIVas

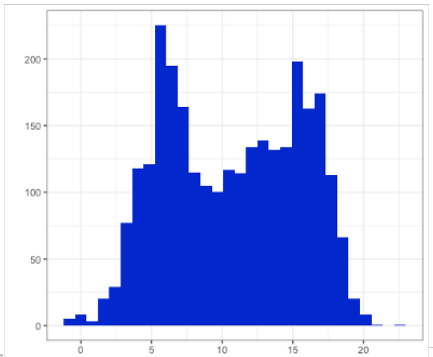
Population density [4]




Office for National Statistics

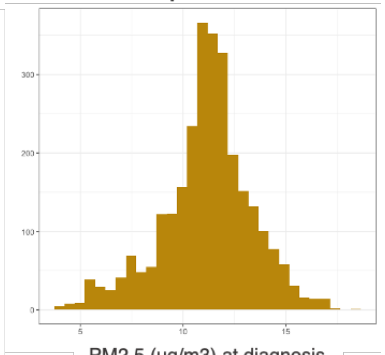
UK Data Service
Census Support

Temperature [5]




Met Office

Air pollution [6]




defra

Department for Environment Food & Rural Affairs

[3] https://ukivas.ndorms.ox.ac.uk/
[4] http://geoconvert.ukdataservice.ac.uk/
[5] https://www.metoffice.gov.uk/research/climate/maps-and-data/data/haduk-grid/datasets
[6] https://uk-air.defra.gov.uk/data/pcm-data

**Table 4.2:** Characteristics of ANCA-associated vasculitis cases from UKIVAS and population controls from UK Biobank

| | UKIVAS AAV cases | | | UK Biobank | p value |
|---|---|---|---|---|---|
| | **MPA** | **GPA** | **EGPA** | **Controls** | |
| n | 757 | 1,612 | 413 | 82,713 | |
| Age, mean (SD) | 65.70 (10.39) | 60.05 (10.50) | 57.43 (10.13) | 59.81 (6.86) | <0.001 |
| Gender | | | | | |
| Female (%) | 405 (53.50) | 761 (47.21) | 207 (50.12) | 48,816 (59.02) | |
| Latitude | | | | | |
| Median (IQR) | 51.87 (51.55, 52.76) | 52.45 (51.66, 53.38) | 52.24 (51.59, 53.39) | 53.36 (51.59, 53.79) | <0.001 |
| Population density | | | | | |
| Urban (%) | 522 (73.5) | 1016 (73.1) | 245 (70.0) | 70726 (86.2) | <0.001 |
| Index of multiple deprivation, n(%) | | | | | |
| 1 (Least deprived) | 130 (20.50%) | 257 (20.49%) | 57 (18.57%) | 14,124 (20.02) | <0.001 |
| 5 (Most deprived) | 122 (19.24%) | 267 (21.29%) | 65 (21.17%) | 14,103 (19.99%) | |
| Ethnicity, n (%) | | | | | 0.003 |
| White British | 606 (80.2) | 1438 (89.3) | 364 (88.1) | 74820 (90.6) | |
| White Irish | 21 (2.8) | 27 (1.7) | <10 (0.5) | 1990 (2.4) | |
| Any Other White Background | 45 (6.0) | 42 (2.6) | 11 (2.7) | 2045 (2.5) | |
| Chinese | <10 (0.7) | <10 (0.1) | <10 (0.2) | 122 (0.1) | |
| Any Other Mixed Background | <10 (0.4) | <10 (0.1) | <10 (0.2) | 117 (0.1) | |
| African | <10 (1.2) | <10 (0.4) | <10 (0.2) | 336 (0.4) | |
| Caribbean | 11 (1.5) | <10 (0.4) | <10 (0.7) | 566 (0.7) | |
| Any Other Asian Background | <10 (0.7) | 10 (0.6) | <10 (0.5) | 192 (0.2) | |
| Bangladeshi | <10 (0.4) | <10 (0.1) | <10 (0.2) | 21 (0.0) | |
| Pakistani | <10 (0.9) | 13 (0.8) | <10 (1.5) | 288 (0.3) | |
| Indian | 19 (2.5) | 37 (2.3) | <10 (1.7) | 831 (1.0) | |

### 4.8.3. Results of the pilot study

Given that this was a feasibility study, the results of the linkage are provided as descriptive summaries and are compared using one way ANOVA for hypothesis testing (Table 4.3). No modelling approach were undertaken as the study was designed to be exploratory and would not generalisable due to differences between UKIVAS and UKB in terms of design and variables collected. Regardless of this, findings from descriptive summary shows that AAV cases from UKIVAS and controls from UKB were similar in terms of age and ethnicity. MPA patients were older compared with controls while GPA and EGPA patients were similar in age with UKB controls. Worth noting is the additional linkage conducted here which led to the addition of small-area data of the UKIVAS registry. The UKIVAS did not collect demographic data related to the patients' place residence, specifically population density and the index of multiple deprivation, which are important confounders in exposure-disease research. These variables were added to the registry using census data from the Office of national statistics (ONS) at a lower super output area (LSOA)[7]. The addition of these two variables revealed that the majority of AAV patients from UKIVAS were from urban areas (72%) with higher proportion of GPA and EGPA patient being from more deprived areas compared with UKB controls and MPA patients compared with UKB study controls (Table 4.2). The linking of environmental data (air pollution data) showed that there were significant differences in air pollution exposure between AAV cases from UKIVAS and UKB controls (Table 4.3). Specifically, AAV patients showed to reside in areas with high levels of particulate matter compared with UKB controls (median $PM_{10}$: 16.6 vs 15.80, $p<0.001$ and $PM_{2.5}$: 11.3 vs 10.7, $p<0.001$). In contrast, exposure to transport related pollutants, like $NO_2$ and NOx was higher in UKB controls compared AAV cases (median $NO_2$: 17.06 vs 20.72, $p<0.001$, NOx: 24.6 vs 30.69, $p<0.001$ ). This difference in nitrogen dioxide and oxide exposure may reflect the fact that UK Biobank is mainly an urban cohort, with 86% of the population being from Urban areas compared with 72% of UKIVAS. Similarly, for $SO_2$ and Benzene,

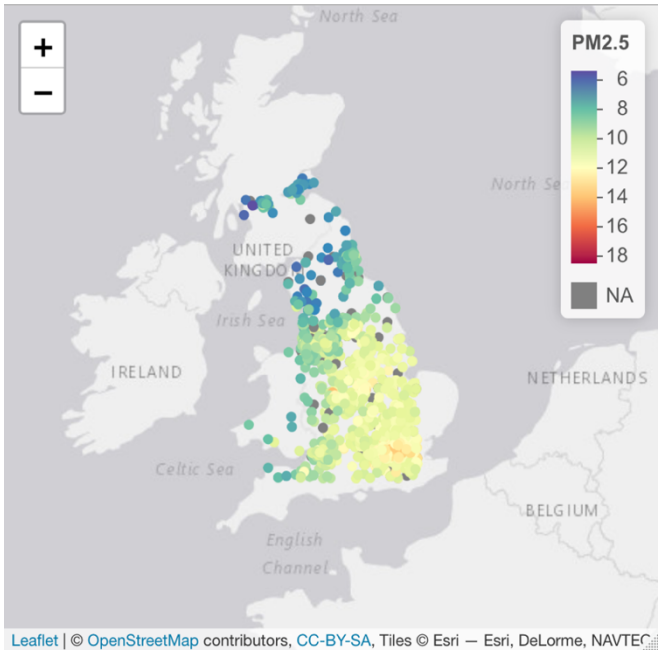
[7] http://geoconvert.ukdataservice.ac.uk/

exposure estimates were higher for UKB controls compared with UKIVAS AAV cases (median $SO_2$: 2.16 vs 2.99, $p<0.001$, Benzene: 0.49 vs 0.55, $p<0.001$)

**Table 4.3:** Summary statistic of residential pollution exposure estimates for ANCA-vasculitis cases and population controls after the linkage.

| | UKIVAS registry | | | UK Biobank | |
|---|---|---|---|---|---|
| | **MPA (n=757)** | **GPA (n=1,612)** | **EGPA (n=413)** | **Control group (n= 82,713)** | **ANOVA p-value** |
| **Pollution exposure, median (IQR)** | | | | | |
| PM10 ($ug/m^3$) 2004-2018 | 16.98 (15.41, 18.64) | 16.57 (14.86, 17.98) | 16.30 (14.23,17.85) | 15.80 (14.15, 17.57) | <0.001 |
| PM2.5 ($ug/m^3$) 2004-2018 | 11.60 (10.61, 12.91) | 11.30 (10.14, 12.31) | 11.07 (9.76, 12.11) | 10.77 (9.57, 12.02) | <0.001 |
| $NO_2$ ($ug/m^3$) 2004-2018 | 17.85 (12.35, 23.77) | 17.77 (12.71, 22.68) | 15.56 (11.68,20.65) | 20.72 (16.30, 25.37) | <0.001 |
| NOx ($ug/m^3$) 2004-2018 | 25.96 (16.91, 37.61) | 25.90 (17.55, 35.44) | 21.94 (15.89,31.15) | 30.69 (22.78, 40.22) | <0.001 |
| $SO_2$ ($ug/m^3$) 2004-2018 | 1.93 (1.40, 2.65) | 2.39 (1.70, 3.29) | 2.17 (1.46, 3.18) | 2.99 (2.23, 3.92) | <0.001 |
| Benzene ($ug/m^3$) 2004-2018 | 0.51 (2.04, 3.29) | 0.50 (2.05, 3.23) | 0.47 (1.93, 2.88) | 0.55 (2.19, 3.65) | <0.001 |

A) ANCA-vasculitis cases




B) Controls

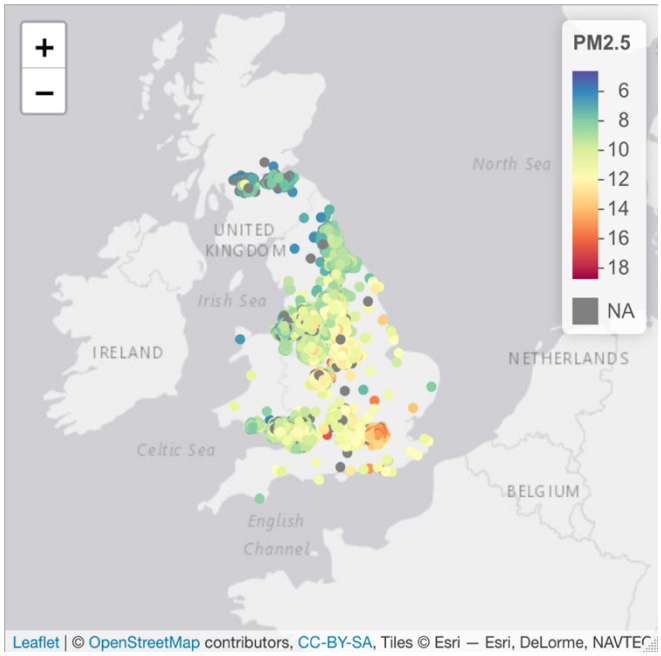
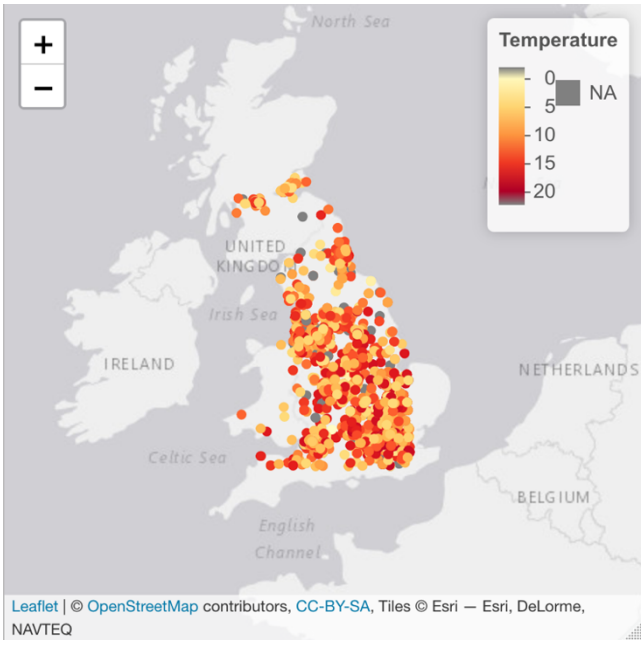



**Figure 4.8:** Schematic summary of the mean PM2.5 exposure estimates in AAV cases (A) and controls (B). These estimates are a 6-year average exposure from the date of diagnosis. Each dot represents an exposure dose for case or control

C) ANCA-vasculitis




D) Controls

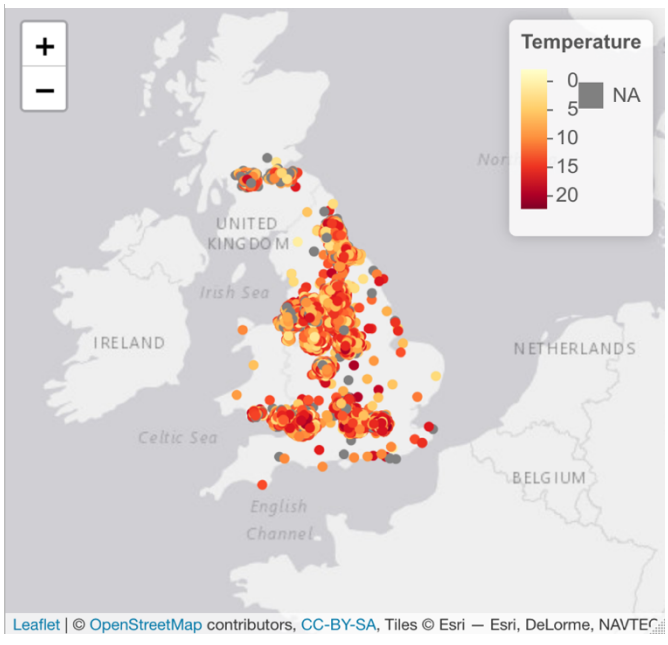



**Figure 4.9:** Schematic summary of temperature exposure estimates of cases (A) and controls (B) at the month of diagnosis

### 4.8.4. Appraisal of the pilot study

#### 4.8.4.1.1. Validity of linking a clinical registry with a biobank

The results from the pilot study demonstrated the possibility of merging two population health registries designed with different purposes to address important questions related to the environmental aetiologies associated with vasculitis. The linking of a clinical registry with population biobank is common in several settings, with cancer registries being seen as a major example. For instance, cancer diagnoses recorded in UK Biobank were linked from the UK national cancer registry which hold records of all people diagnosed with cancer each year, their treatment and prognostic outcomes (287). Linking this data with the UKB and its rich biomarker samples and environmental data, have allowed for important research into new aetiologies associated with cancer. Examples of these can be seen in recent nested case control publications showing significant association between air pollution exposures and an increased risk in lung and laryngeal cancer (288–290).

Similar linkages have also been reported in several biobanks from Nordic countries (Finland, Sweden, and Norway) (287,291). In these examples, direct linkage of cancer registries to biobanks are achieved using unique set of personal identifiers that are precise, robust and stable over time (292). This form of linkage known as exact match linkage is regarded as gold standard when linking multiple health registries. In some cases, indirect linkages using a mix of numerous variables at a group level are used. Kollhorst et al.,2021 recently demonstrated possible linkage approach using indirect personal identifiers, year of birth, sex, area of residence, and date of diagnosis, to merge insurance claim data with cancer registry data (293). Such indirect linkage approach, though not as sensitive and specific as exact matching, has the potential to inform aetiology research in registries with a lack of access to control groups. Such linkage approaches may have utility in informing future linkages of rare disease registries that can facilitate research of potential risk factors of vasculitis.

### 4.8.4.2.Generalisability of pilot study

The generalisability of research studies assessing the effect of environmental exposure on health is dependent on the design (e.g.: sampling strategy) and accuracy of linkage features such as the timing and place of the health event (e.g.: time of diagnosing and patient address). Care should be taken at all stage of design and planning to minimise threats to validity, mostly selection bias, measurement error, or missing data. For this pilot study, the validity of the results from the exploratory analyses reflects the non-random nature of the sampling approach used to identify cases and controls as well as major differences in sampling strategy between UKIVAS and UK Biobank. Specifically, UK Biobank sampling strategy targeted 22 regions of the UK with the aim to have a representative sample of the UK population (between 37 – 70 years old) and in term of rural and urban classification (e.g.: urban population of UKB was 86.5% vs 82.4% urban population from the 2011 census)(294). For UKIVAS, the sampling strategy intended to cover as much of the UK population and to reflect the vasculitis population as seen with higher proportion of UKIVAS population being from rural areas compared with UKB and UK census populations (UKIVAS rural population: 28.8% vs 13.52% of UK Biobank and 17.6% 2011 census). This difference in rural and urban population was reflected in the air pollution exposure measures between cases and controls (Appendix 13). Controls from UKB had higher exposure to urban pollutants like $NO_2$ and NOx. Similar differences in particulate matter ($PM_{10}$ and $PM_{2.5}$) were seen and may also reflect the sampling approach between the two 'registries. These considerations and limitation were used to inform analyses in chapter 5 and 6 and are further discussed in the section below

### 4.8.4.3.Key considerations for this thesis

As noted earlier in chapter 2 and now with the pilot study, the external validity of studies investigating the effect of environmental exposure on vasculitis are mainly affected by sampling approach and the source health data used. Selection of cases and controls should consider the use of random sampling techniques to minimise systematic differences that may be due to selection

bias. These consideration are accounted for in the upcoming result chapter which will be the first studies to use data from a biobank and administrative health data to investigate environmental risk factors associated with vasculitis.

## 4.9. Summary

The first part of this chapter introduced the history of record linkage of routine health data. It provided a summary of important events that led the evolution of record keeping method that led birth of routine data which can now be linked with other dataset such the census and environmental data to investigate social and environmental determinants of health. The benefits and limitation of routine using routine data for research were also discussed. This included the ability to do cost-efficient research with large sample size studies that cover hard-to-reach populations. Limitations related to data pre-processing, inconsistency and missingness of data were also discussed. A detailed summary of how routine data is used to investigate the impact of environmental exposures on health were provided, together with associated linkage methods, from the sampling strategy to linkage approaches commonly used to conduct environmental-wide associated studies. The last part of this chapter provided a demonstration of application of these methods in a pilot study linking air pollution data to UKIVAS. Results of the pilot study informed linkage approaches and analyses presented in the results section

# Chapter 5: Understanding the long-term impact of residential air pollution exposures on vasculitis and other autoimmune diseases in UK Biobank

## 5.1. Overview

This chapter will report the first result of a study showing the long-term effects air pollution exposure on the risk of vasculitis an EWAS study approach. The term systemic vasculitis will be used to refer to patients from UKB who were diagnosed with a small to large vasculitis based on a three-character ICD coding for vasculitis (M30-31). These cases were identified from the first occurrence data fields described in Chapter 3. Eligibility criteria used to guide selection of vasculitis cases and disease comparators (rheumatoid arthritism, systemic lupus erythematosus, ankylosing spondylitis and psoriatic arthritis) will be provided. The inclusion of these inflammatory autoimmune disease comparators mentioned in Chapter 2 were based on the evidence of shared genetic risk profiles, including pathogenic pathways and potential environmental triggers (137,295–300). For instance, results from the review in chapter 2 showed that the effects of air pollution on vasculitis varied based on geography and method of exposure assessment including linkage approach undertaken(216). Similar heterogeneities are seen in studies reporting effects of air pollution on inflammatory autoimmune diseases like rheumatoid arthritis and systemic lupus erythematosus. Some studies report protective effects of air pollution (like particulate matters, $PM_{10}$ and $PM_{2.5}$) and others have reported adverse effects (like Ozone, $O_3$)(198).

Comparing the impact of air pollution on vasculitis or any other autoimmune disease is difficult because of the general differences in population source and method of recruitment, especially sampling strategy and selection of controls, including tools for exposure assessment and adjustment of confounders. Using UK Biobank provides opportunity to overcome these limitation. This chapter will demonstrate a comparison study of effects of air pollution on vasculitis and

other autoimmune diseases in this large population cohort that may be generalisable to the vasculitis population in the UK. A variety of potential confounding measures will be considered as well as key environmental linkage methods discussed in chapter 4. Additionally, this chapter will take advantage of UK Biobank rich biomarker data collected at baseline to investigate short-term effects of air pollution on surrogate markers of vasculitis and other major autoimmune diseases. This is an important addition because smaller observational studies have shown that air pollution exposures can induce short-term increases in important markers of inflammation and immunity, including a general rise in white blood cell count, C-reactive protein, and fibrinogen(74,301–304).

## 5.2. Method

### 5.2.1. Participants

This chapter reports on analyses from the baseline and follow-up data from participants recruited in UK Biobank (19). As highlighted in chapter 3, UK Biobank enrolled over a half-million adults, aged 37 – 73 years, between 2006 and 2010. Individuals who resided within 25 miles of the UKB study centres and who were registered with the National Health Service (NHS) were invited to take part in the study. Participants were enrolled across 22 centres in the UK, covering a variety of settings, in terms of sociodemographic heterogeneity and urban-rural mix. At baseline, participants provided detailed personal, demographic, and health-related information via a touchscreen questionnaire and nurse interview. Additional information about individuals' current and historic health outcomes were retrieved from the NHS central repository as well as various national health registries. This data was linked centrally by UK Biobank to each participant following baseline assessment. Everyone enrolled in the study gave a written informed consent. This chapter and the following analyses were conducted under the generic ethical approval from the UK Biobank (REC: 21/NW/0157 ).

### 5.2.2. Outcome measures

#### 5.2.2.1. Defining systemic vasculitis and other autoimmune diseases

Systemic vasculitis and other autoimmune disease were selected and defined using the International Classification of Disease (ICD10 and ICD9) chapters (see Appendix 11 for complete of included ICD codes) . Participants who were diagnosed or who reported having a systemic vasculitis (ICDM30-31), rheumatoid arthritis (ICDM05-06), systemic lupus erythematosus (ICDM32), ankylosing spondylitis (ICDM45) and psoriatic arthritis (ICDM07) were screened and included in the final analysis of this chapter. Those assigned more than one of autoimmune disease diagnoses simultaneously at their last date of diagnosis was excluded (n=86) (Figure 1). As highlighted in previous paragraph, health outcome data and other clinical information pertaining to each study participants were retrieved by means of record linkage of routine data from the NHS. This linkage was conducted centrally by the UK Biobank team (https://biobank.ndph.ox.ac.uk/ukb/refer.cgi?id=149480).These sources of outcome data included hospital inpatient records, the primary care records, the cancer registry, and death records and self-reported conditions captured at baseline by each participant. The selection and identification of cases used in the chapter were captured under the first occurrences data fields which were generate by specially designed algorithm mapped using the International Classification of Disease (ICD10 and ICD9) chapters (https://biobank.ndph.ox.ac.uk/ukb/refer.cgi?id=2112) .

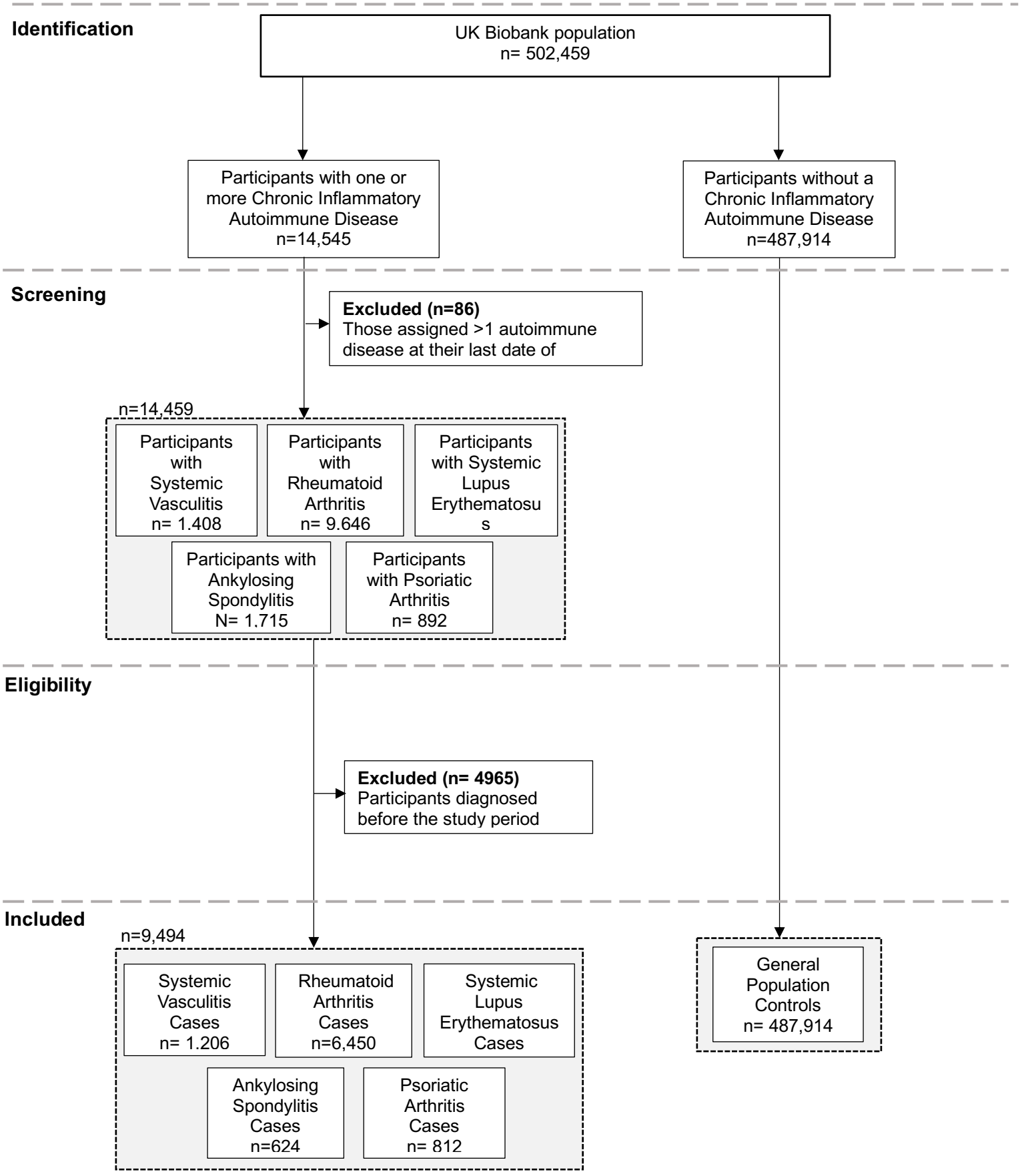



**Figure 5.1:** Study Flowchart summarising the eligibility criteria of participants included in the final analysis of this chapter.

### 5.2.2.2. Peripheral markers of disease activity and inflammation

#### 5.2.2.2.1. Blood sampling

Each participant enrolled in the UK Biobank had their blood collected at baseline using a standardized and research-validated protocol for blood collection. In summary, blood samples were drawn by a phlebotomist, or a nurse trained and certified to conduct blood collection. For collection, each participant was seated in a curtain blood collection office and was asked for about 50ml of blood sample. This was the final stage of the baseline assessment. Following blood collection, samples were packed and transported according to an established protocol and were processed within few hours (<24hrs) of collection.

#### 5.2.2.2.2. C- reactive protein and rheumatoid factor

C-reactive protein (CRP) and rheumatoid factor (RF) were generated and quantified using an immunoturbidimetric method (Beckman Coulter AU5800 immunoassay analyser). This is classical antigen-antibody reaction approach which generate particles that can be optically detected by a photometer. Measurements from the CRP assay were verified against the manufacturer's performance standards. Linearity and limit of detection determined the observed reportable range. Samples exceeding this range were diluted and re-analysed. Quantitative measures captured from the CRP and RF assay are reported in mg/ml.

#### 5.2.2.2.3. Neutrophil, eosinophil, and monocyte count

Neutrophil, eosinophil, and monocyte count were generated using a Beckman Coulter LH750 instrument. Results of neutrophil, eosinophil and monocyte count are provided as a proportion (e.g.: neutrophils/100) x total white blood cell count.

## 5.3. Exposure measures

### 5.3.1. Air pollution measures

The pollution measures used in this chapter were sourced from the department for Environment, Food and Rural Affairs (DEFRA) and were linked to UK Biobank dataset by the thesis author. These measure include particulate matter with aerodynamic diameter of up to 10um and 2.5um ($PM_{10}$ and $PM_{2.5}$), nitrogen dioxide ($NO_2$), nitrogen oxide ($NO_x$), sulphur dioxide ($SO_2$) and benzene ($C_6H_6$). These exposures represent outdoor background concentrations modelled at 1kmx1km resolution using a combination of historic and recent statistical modelling that cover a variety of emission sources and land surface for the period of 2001 to 2018 (305–307). These exposure estimates were derived from the pollution climate mapping (PCM) models (https://uk-air.defra.gov.uk/data/pcm-data) consisting of:

i) Air dispersion models and emission estimates from the UK National Atmospheric Emissions Inventory, for large point source (e.g.: a power station) - $NO_x$, $SO_2$, $PM_{10}$ and $PM_{2.5}$.

ii) Dispersion and linear models and repeated measurements collected from small point sources (e.g.: factories) - $NO_x$, $SO_2$, $PM_{10}$ and $PM_{2.5}$

iii) Modelled estimates from distant sources, mainly rural background sources - $NO_x$, $SO_2$, $PM_{10}$ and $PM_{2.5}$.

iv) Area sources related to domestic combustion, this being defined as diffuse emissions from unspecified locations (e.g.: domestic heating or emissions from shipping), small point sources related to combustion industry, road traffic and other area sources. For these areas, pollution estimates were modelled using time varying kernel dispersion models and emission estimates from the national atmospheric inventory (2001-2018) - $NO_x$, $SO_2$, $PM_{10}$ and $PM_{2.5}$

v) Fugitive source kernel modelling, estimating emission from fugitive components derived from the National Atmospheric Emission Inventory (2001-2018)- $NO_x$, $SO_2$, $PM_{10}$ and $PM_{2.5}$.

vi) Interpolated and dispersion (NAME) models of secondary organic and inorganic aerosols formed by oxidation of non-methane volatile organic

compounds and scaled measurements of $SO_4$, $NO_3$, $NH_4$ at rural sites - for $PM_{10}$ and $PM_{2.5}$.

vii) Dispersion kernel models, vehicle activity, and land use data on regional calcium rich dusts from re-suspended soil, iron rich dusts from heavy duty vehicles - $PM_{10}$ and $PM_{2.5}$

viii) Interpolated and scaled measurements of chloride and sea salt at rural sites - $PM_{10}$ and $PM_{2.5}$

ix) Oxidant-partitioning model which describes the complex interaction and relationship between NO, $NO_2$ and $O_3$ (Jenkin, 2004; Murrells et al., 2008; Jenkin, 2012)

The decision to include above pollutants is based on a pragmatic approach; firstly, because currently there is no evidence to support any association between air pollution and systemic vasculitis. Secondly, there is a need for a broad understanding about the long-term impact of different air pollutants on vasculitis onset, and potentially relapse. Second rationale is that for pollutants such as $PM_{10}$, $PM_{2.5}$ and $SO_2$, there is a large body of evidence to show they are able impact health and are associated with adverse outcomes, including cardiovascular and respiratory diseases including cancer.

### 5.3.2. Data linkage

UK Biobank provided data on participants residential history including changes in participants address. This data was sourced from the Administrative Data Liaison Service (ADLS) via GP registration and is managed by the NHS digital for participants residing in England and Wales. For Scotland, the same data were provided by the Information Services Division (ISD). Where full address was present and verified to be valid, the software package, DataPlus, was used to transform and geocode addresses into grid coordinates. Where only the postcode was present, the coordinates were generated using Doogal (https://www.doogal.co.uk) and UK postcodes (https://www.uk-postcodes.com). The grid coordinates are provided in the British National Grid (OSB1936) projection and refer to easting and northing with a reference point near the Isles of Scilly. UK Biobank provided these coordinates at 1km grid

postings. These coordinates and past coordinates corresponding to study participants address history between 2001 and 2018 were link residential air pollution exposures. For the data linkage, given the pollution data were provided on a 1kmx1km grid square maps with associated central point (centroid), pollution values were calculated and assigned based on intermediate zone around a participant centroid. Specifically, participants who resided within a particular intermediate zone were assigned a mean value for that zone. For zones where there were no grid square centroids, pollution estimates from the nearest grid square was used. It is worth noting 55% of UK Biobank population who did not have a primary care incident, or a hospital admission did not have longitudinal information about their address changes. A median air pollution exposure covering the period between 2001 to 2018 was assigned to each participant, both cases and controls.

## 5.4. Other measures

### 5.4.1. Socio-demographic

Participants age was derived by UK Biobank from individual's date of birth and the first date of visit to the assessment centre and is recorded in whole years. Gender was self-reported as male or female and in some cases was acquired from the NHS central registry at recruitment. Ethnicity was self-reported and is recorded categorically as either White, Asian/Asian British, Black/Black British, Chinese, or mixed/other ethnic group. Education was self-reported and for this study, it is dichotomous so to look at participants with and without a university degree. Smoking status was self-reported and was recorded categorically as never, previous, and current smoker. The frequency of alcohol intake was self-reported and is recorded categorically as daily/almost daily, three to four times a week, one to two times a week, one to three times a month, special occasion only or never.

Townsend deprivation index was calculated by UK Biobank immediately prior to participants joining the UK Biobank. This was based on the preceding national census output areas. Each participant was assigned a score

corresponding to their output area in which their postcode is located. For this study, the Townsend deprivation score was transformed into quintiles, with the first category being representative of individuals from least deprived output areas and the last category being representative of the most deprived areas.

Physical activity measure was generated using a number of self-reported factors (e.g.: walking, moderate and vigorous physical activity, strenuous sports…etc) related to the type and duration of physical activity that an individual undertook in their daily life (239). These factors were converted into a single measure of total physical activity in metabolic equivalent of task (MET) – hours per week weighted by intensity (walking, moderate or vigorous). Lastly, self-reported time spent outdoors was derived based on the average time (in hours) that participants spent outdoors both in winter and summer.

### 5.4.2. Other environmental measures

Home area population density was derived by UK Biobank by combining each participant postcode with data from 2001 census from the office of National Statistics 2001. using the Geoconvert tool (http://geoconvert.mimas.ac.uk/ ) from the Census Dissemination Unit. This was measured categorically as Urban, Town and Fringe, Village, Hamlet and isolated dwelling and rural mix. Close to major road was recorded as binary indicating whether a residential address was within 50 metres of a class 1 or 2 type of road (See **Appendix 14** for detail summary of variables used in this chapter and how they were coded)

## 5.5. Statistical analysis

Statistical analyses were performed in R studio v1.2.5019 and Stata v13. Baseline characteristic of participants with vasculitis and other autoimmune diseases were summarised assessed using descriptive statistics. For normally distributed variables (age), mean and standard deviation is reported. For variables with skewed distribution (physical activity, time spent outdoors and pollution), median and interquartile range were calculated and is reported.

Relative frequencies in percentage are reported for categorical data. Incidence rate was calculated as the sum of all new cases of vasculitis (n=802) captured between the start of UKB recruitment and the end follow up period chosen for this chapter (2006 and 2018) divided by the size of UKB cohort (n= 502,459) then separately by UK population estimates of people between the age 40 – 75 years old (n = 25.1 million from 2011 mid-term population estimates). To assess the effects of air pollution on vasculitis and other autoimmune diseases, logistic regression was used. All the outcomes reported here were modelled separately (1-cases vs 0- population control), first unadjusted and adjusted for potential confounders (age, gender, education, ethnicity, Townsend deprivation score, smoking and alcohol intake status and time spent outdoors and population density). All covariables included in the modelling stage were within reasonable range of multicollinearity (variation inflation factor (VIF) =1.7 to 1.9) and were selected to minimise confounding. Given there was a total of 30 cross-sectional analyses combined for the outcomes of interest, the two-tail p-values were adjusted using Simes-Benjamini-Hochberg false discovery rate (FDR) method(308), both for the unadjusted and adjusted. Additionally, a stratified analysis was undertaken to get insight into population groups that are at a risk in terms of geography (urban vs rural), deprivation (least vs most deprived), lifestyle (smokers vs non-smokers), and time spent outdoors. The stratified analyses involved fitting separate models (unadjusted and adjusted for potential confounders) for pollutants that showed to be significantly associated with vasculitis or other autoimmune diseases. Stratified models were based on variables population density, deprivation, smoking, and time spent outdoors. For markers of inflammations, a linear regression was used to assess the association between annual mean pollution exposures at baseline and log transformed concentration of CRP and rheumatoid factor, including neutrophil, eosinophil, and monocyte count. The standardised/beta coefficients result were multiplied by 100 to give percentage of log transformed increase/decrease in general biomarkers of inflammation. These results are presented using a feature expression heatmap (309)

### 5.5.1. Sensitivity analysis

To assess the robustness of the results and the extent to which the point estimates could be affected by changes in methodology and study sample size, a time to event analysis was undertaken, restricted to participants who entered UKB without the disease. Specifically, the analysis included vasculitis and other autoimmune disease cases who were diagnosed after baseline (date of attending assessment centre). The follow up was calculated as time from date of assessment centre visit to the date of disease of diagnosis (date of first occurrences) or death if the participant died before the disease onset. A cox proportional hazard model was used, first as unadjusted and adjusted for age, gender, education, ethnicity, Townsend deprivation score, smoking and alcohol intake status and time spent outdoors and population density.

## 5.6. Results

### 5.6.1. Cohort characteristics

**Table 1** provides a detailed summary of cohort characteristics and median pollution exposure in vasculitis cases and other autoimmune diseases including the rest of the UK Biobank population. From the full cohort (n=502,459), 14,460 participants were identified to have one or more inflammatory autoimmune disease. From this group, 9,494 had all the pollution exposure measures linked to their place of residence and were eligible for inclusion in the final analysis. This included 1,206 participants with systemic vasculitis, 6,451 with rheumatoid arthritis, 401 with systemic lupus erythematosus (SLE, ), 624 with ankylosing spondylitis (AS), 812 with psoriatic arthritis (PsA).

The incidence of vasculitis was 31.95 cases per million population when calculated using 2011 census population estimates (n=25.1 per million between 40 and 75 years old). This incidence is higher than normal when compared against studies reported in Chapter 1 (incidence range: 15.1 to 23.1 cases per million population 50 years and older). Additionally, the incidence rate calculated using UKB complete cohort (n=502,459) was 16.45 (95% CI: 15.36

– 17.64) cases per 100,000 person-years. This incidence rate was particularly higher in females, smokers, and those without a university degree (**Appendix 15**).

As noted in chapter 4, the majority of UKB participants were from urban areas, reflecting UKB recruitment approach (**Figure 2**). When compared with the UK Biobank control group (n=487,942), a higher proportion of cases with vasculitis were females (64% vs 54%), without a university degree (76% vs 67%), from the most deprived areas (22% vs 19%) and generally less active (24 vs 28 METhrs/week). Additionally, vasculitis cases were older compared with other autoimmune diseases participants (ANOVA p<0.001).

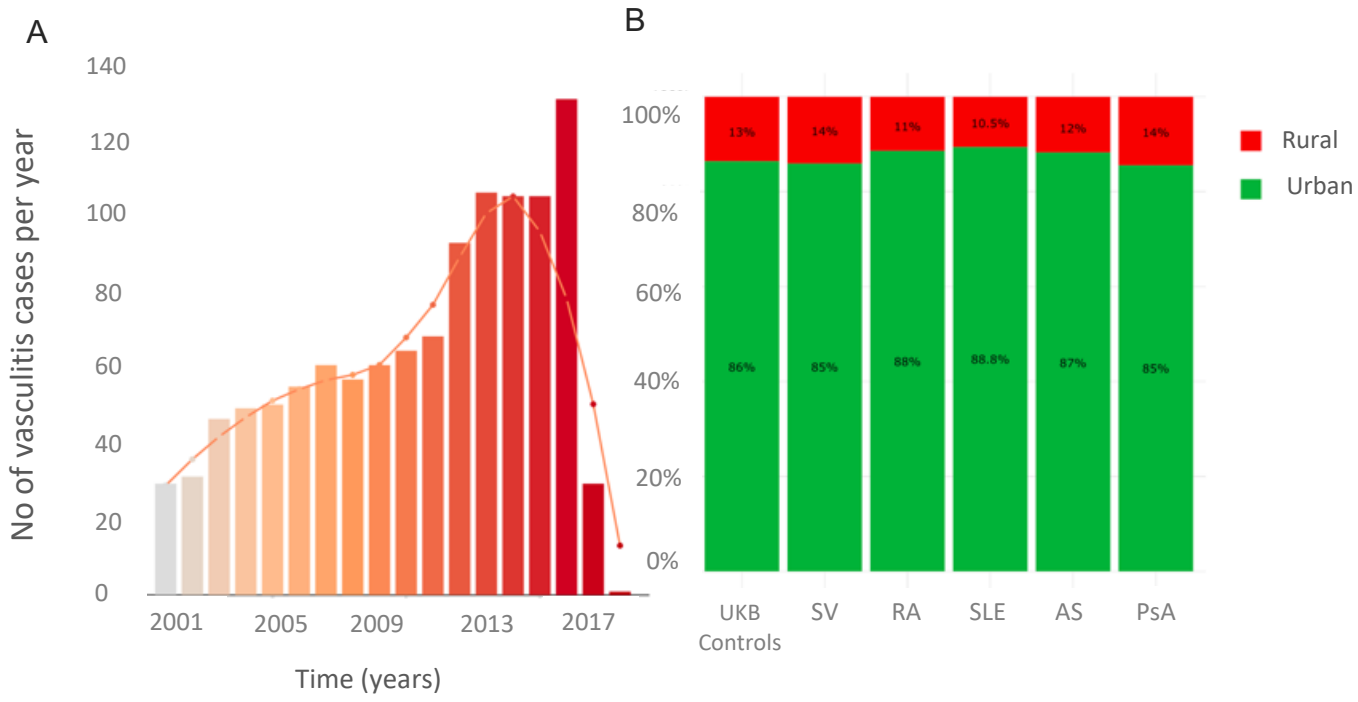



**Figure 5.2:** Number of systemic vasculitis cases diagnosed during the study period (2001-2018) in UK Biobank (A) and the urban and rural distribution of these participants together with other autoimmune diseases and the rest of the UK Biobank cohort (B)

Besides age, cases with vasculitis were relatively similar with other autoimmune disease participants in terms of education and behavioural risk factors like smoking and alcohol intake. Interestingly, both vasculitis and other autoimmune disease participants were from the most deprived areas (26% vs 19.8)

compared with UKB population ($X^2$p-value=0.001). Those with rheumatoid arthritis and ankylosing spondylitis lived in areas near major roads compared with UKB controls. With regards to air pollution measures, there were general differences in air pollution exposure between vasculitis cases and controls (**Appendix 16**). Similar differences were also seen for other autoimmune disease when compare with UKB controls.

**Table 5.1:** UKB Cohort characteristics of people with systemic vasculitis and other autoimmune disease and the general population controls

| | **Systemic Vasculitis** ***n* = 1,206** | **Rheumatoid Arthritis** ***n* = 6,451** | **Systemic Lupus Erythematosus** ***n* = 401** | **Ankylosing Spondylitis** ***n* = 624** | **Psoriatic Arthritis** **n=812** | **Population Controls** ***n* = 487,942** |
|---|---|---|---|---|---|---|
| **Age** mean (±SD), years | 61.41 (±6.33) | 58.99 (±7.30) | 55.94 (±8.34) | 57.42 (±8.07) | 56.85 (±7.56) | 56.46 (±8.11) |
| **Gender** female, *n* (%) | 759 (62.9) | 4342 (67.3) | 334 (83.3) | 310 (49.7) | 450 (55.4) | 264,026 (54.1) |
| **Ethnicity,** *n* (%) | | | | | | |
| White | 1,138 (94.8) | 6,002 (93.7) | 349 (88.1) | 589 (95.0) | 779 (96.4) | 459,032 (94.6) |
| Asian or Asian British | 24 (2.0) | 173 (2.7) | 11 (2.8) | 11 (1.8) | 14 (1.7) | 95,57 (2.0) |
| Black or Black British | 15 (1.2) | 123 (1.9) | 25 (6.3) | 6 (1.0) | 1 (0.1) | 78,32 (1.6) |
| Chinese | 3 (0.2) | 18 (0.3) | 1 (0.3) | 1 (0.2) | 3 (0.4) | 1,542 (0.3) |
| Mixed or other ethnic groups | 20 (1.7) | 87 (1.4) | 10 (2.5) | 13 (2.1) | 11 (1.4) | 7,302 (1.5) |
| **Educated to degree level** | | | | | | |
| Yes, n (%) | 285 (24.3) | 1380 (21.9) | 110 (28.6) | 148 (24.3) | 212 (26.5) | 157,774 (33.0) |
| **Townsend score quintile** | | | | | | |
| 1 Least deprived, *n* (%) | 238 (19.8) | 1106 (17.2) | 57 (14.2) | 95 (15.2) | 145 (17.9) | 98,104 (20.1) |
| 5 Most deprived | 269 (22.4) | 1690 (26.2) | 118 (29.4) | 176 (28.2) | 198 (24.5) | 96,713 (19.8) |
| **Total Physical activity (METhrs/week)** | | | | | | |
| Median (p25, p75) | 23.97 (10.55,50.93) | 24.50 (9.95, 56.00) | 23.10 (10.48, 48.20) | 27.30 (11.55,57.31) | 19.93 (8.93, 49.32) | 28.00 (12.60,57.70) |
| **Smoking status** | | | | | | |
| Never, n (%) | 601 (50.3) | 2931 (45.9) | 207 (52.0) | 290 (46.8) | 376 (46.7) | 266,704 (55.0) |
| Previous | 453 (37.9) | 2557 (40.0) | 147 (36.9) | 241 (38.9) | 336 (41.7) | 167,306 (34.5) |

| | | | | | | |
|---|---|---|---|---|---|---|
| Current | 141 (11.8) | 904 (14.1) | 44 (11.1) | 88 (14.2) | 94 (11.7) | 51,110 (10.5) |
| **Alcohol Frequency** | | | | | | |
| Never, n (%) | 75 (6.2) | 434 (6.8) | 34 (8.6) | 35 (5.6) | 41 (5.1) | 21,433 (4.4) |
| Previous | 72 (6.0) | 459 (7.1) | 26 (6.5) | 38 (6.1) | 70 (8.7) | 17,088 (3.5) |
| Current | 1056 (87.8) | 5528 (86.1) | 337 (84.9) | 548 (88.2) | 698 (86.3) | 447,820 (92.1) |
| **Time Spent Outdoors (in hrs)** | | | | | | |
| Median, (p25, p75) | 2.50 | 2.50 | 2.00 | 2.50 | 2.50 | 2.50 |
| | (1.50, 4.00) | (1.50, 4.00) | (1.25, 3.00) | (1.75, 4.00) | (1.50, 4.00) | (1.50, 3.75) |
| **Close to major road (2008)** | | | | | | |
| Yes, n (%) | 76 (6.4) | 494 (7.8) | 27 (6.8) | 48 (7.8) | 51 (6.3) | 34,304 (7.1) |
| **Traffic intensity on nearest major road (2008) (vehicles/24 h)** | | | | | | |
| Median | 17,353 | 17,139 | 18,964 | 17,674.00 | 17,328 | 17,103 |
| (p25; p75) | (12,98,25,59) | (12,66, 25,31) | (13,73, 27,28) | (13,06,25,81) | (12,42,25,35) | (12,59,25,26) |
| **PM10 ($_{2001-2018}$) (µg/m3)** | | | | | | |
| Median | 15.27 | 15.36 | 15.44 | 15.49 | 15.00 | 15.50 |
| (p25; p75) | (13.62,16.98) | (13.80, 16.93) | (13.97, 17.28) | (13.98,17.03) | (13.57,16.62) | (13.81,17.07) |
| **PM2.5 $_{(2002-2018)}$ (µg/m$^3$)** | | | | | | |
| Median | 10.31 | 10.40 | 10.52 | 10.44 | 10.16 | 10.52 |
| (p25; p75) | (9.16, 11.61) | (9.26, 11.59) | (9.48, 11.84) | (9.37, 11.59) | (9.06, 11.31) | (9.29, 11.74) |
| | | | | | | |
| **NO2 $_{(2002-2018)}$ (µg/m$^3$)** | | | | | | |
| Median | 19.23 | 19.88 | 20.30 | 19.96 | 19.42 | 19.65 |
| (p25; p75) | (15.30,23.11) | (16.13, 23.57) | (16.61, 24.27) | (16.56,23.71) | (15.89,22.78) | (15.76,23.49) |
| **NOx $_{(2002-2018)}$ (µg/m$^3$)** | | | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| Median | 28.73 | 29.89 | 31.08 | 30.25 | 29.22 | 29.48 |
| (p25; p75) | (21.71,36.22) | (23.03, 37.05) | (23.93, 38.28) | (23.82,37.54) | (22.64,35.82) | (22.39,37.03) |
| **SO2 (2002-2018) (µg/m$^3$)** | | | | | | |
| Median | 3.83 | 3.94 | 4.02 | 4.03 | 3.82 | 3.81 |
| (p25; p75) | (3.13, 4.64) | (3.22, 4.69) | (3.22, 4.76) | (3.31, 4.80) | (3.15, 4.66) | (3.10, 4.60) |
| **Benzene (2003-2018) (µg/m$^3$)** | | | | | | |
| Median | 1.60 | 1.65 | 1.69 | 1.66 | 1.59 | 1.62 |
| (p25; p75) | (1.24, 2.01) | (1.31, 2.03) | (1.32, 2.16) | (1.33, 2.01) | (1.27,1.95) | (1.27, 2.02) |

### 5.6.2. Association between long-term air pollution exposure and systemic vasculitis risk and risk for other autoimmune diseases

Figure 5.3 shows the results from 30 separate logistic regression models depicting the long-term the relationship between long-term air pollution exposure and systemic vasculitis and other autoimmune disease. In summary, it was observed that sulphur dioxide ($SO_2$) was adversely associated with systemic vasculitis. This was true before and after adjusting for select confounders. Similar associations were seen in other autoimmune diseases like rheumatoid arthritis and systemic lupus erythematosus. Specifically, a 1 µg per cubic meter higher average level of sulphur dioxide between 2001 to 2018 was associated with 6.4% (adjusted odds ratio: 1.064, 95% CI: 1.004 – 1.127) and 3.7% (adjusted odds ratio: 1.037 95% CI: 1.011 - 1.064) increased odds of systemic vasculitis and rheumatoid arthritis. For systemic lupus erythematosus, 1 µg per cubic meter higher average level sulphur dioxide was associated with 9.7% (adjusted odds ratio: 1.097 95% CI: - 0.994 - 1.211) increased odds of SLE. However, this was not statistically significant as confidence intervals crossed the null. The effects air pollution on rheumatoid arthritis remained statistically significant after adjusting for multiple comparison, but not for systemic vasculitis. (**Table 5.2**). Conversely, particulate matters ($PM_{10}$ and $PM_{2.5}$) were inversely associated with rheumatoid arthritis ($PM_{10}$ adjusted odds ratio 0.965, 95%CI 0.954– 0.978, $PM_{2.5}$ odds ratio 0.947, 95% CI 0.932– 0.964), Ankylosing Spondylitis ($PM_{10}$ adjusted odds ratio 0.954, 95%CI 0.917– 0.994, $PM_{2.5}$ odds ratio 0.929, 95% CI 0.880– 0.981) and psoriatic arthritis ($PM_{10}$ adjusted odds ratio 0.913, 95%CI 0.881– 0.946, $PM_{2.5}$ odds ratio 0.879, 95% CI 0.838 – 0.923), both before and after adjusting for potential confounders.

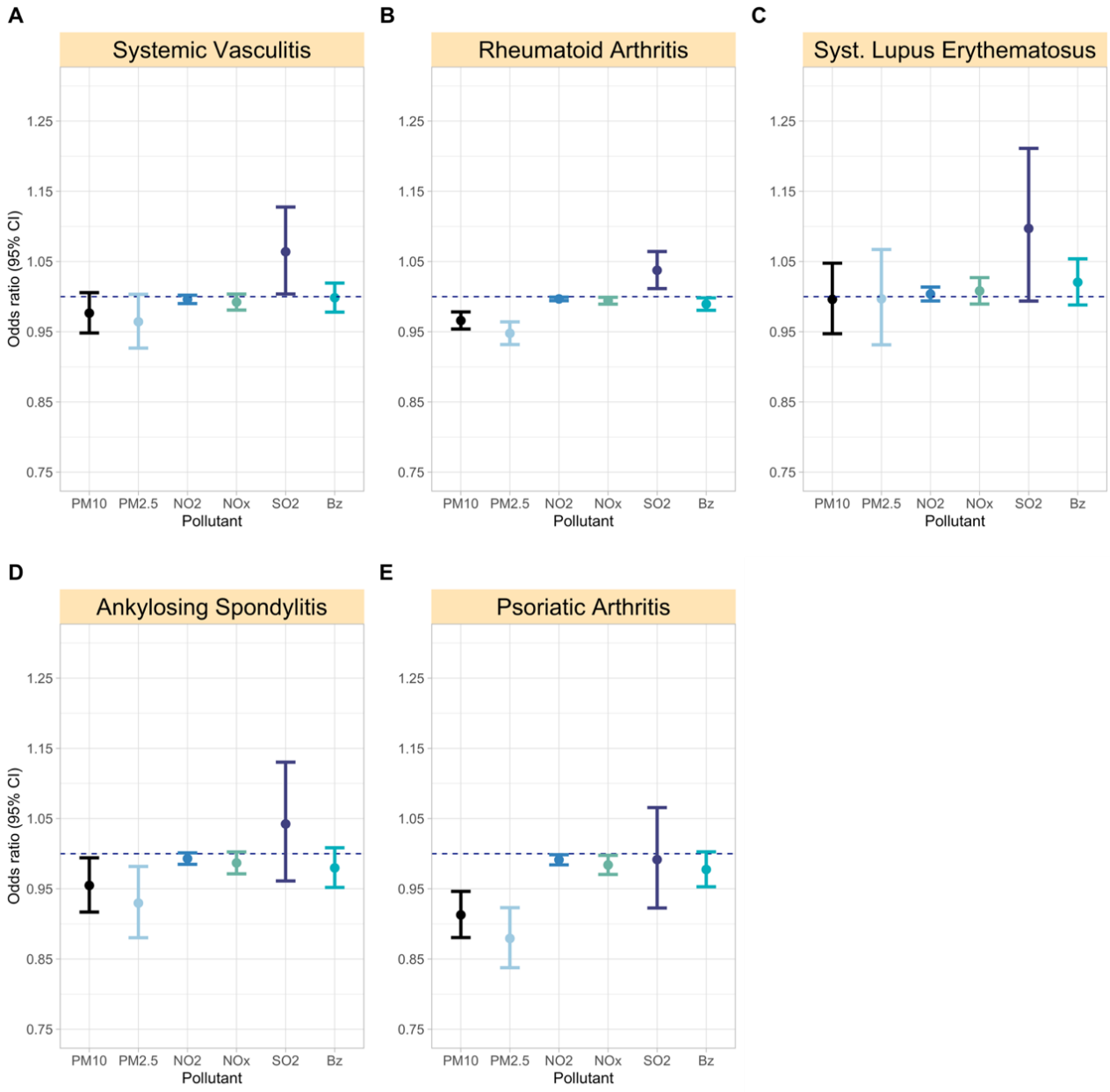



**Figure 5.3:** Odds ratio plot depicting the association between long-term exposure to air pollution (2001-2018) and the odds of systemic vasculitis (A) and other autoimmune diseases (B-E) in UK Biobank

**Table 5.2:** Result of showing the relationship between long-term air pollution exposures and the onset systemic vasculitis and other autoimmune disease in UKB.

| | Pollutant | Unadjusted | | | Adjusted [a] | | |
|---|---|---|---|---|---|---|---|
| | | OR | p-value | p(FDR) | OR | p-value | p(FDR) |
| Systemic Vasculitis | PM10 | 0.974 | 0.029 | 0.073 | 0.977 | 0.115 | 0.182 |
| | PM2.5 | 0.9638 | 0.019 | 0.052 | 0.964 | 0.072 | 0.144 |
| | NO2 | 0.996 | 0.132 | 0.240 | 0.992 | 0.181 | 0.259 |
| | NOx | 0.993 | 0.159 | 0.240 | 0.996 | 0.198 | 0.270 |
| | SO2 | 1.037 | 0.151 | 0.240 | 1.064 | 0.037 | 0.085 |
| | Benzene | 0.996 | 0.695 | 0.719 | 0.998 | 0.882 | 0.912 |
| | | | | | | | |
| Rheumatoid Arthritis | PM10 | 0.982 | 0.001 | 0.004 | 0.966 | 0.000 | 0.000 |
| | PM2.5 | 0.972 | 0.000 | 0.000 | 0.948 | 0.000 | 0.000 |
| | NO2 | 1.001 | 0.103 | 0.238 | 0.994 | 0.020 | 0.060 |
| | NOx | 1.003 | 0.172 | 0.246 | 0.997 | 0.012 | 0.051 |
| | SO2 | 1.083 | 0.000 | 0.000 | 1.037 | 0.004 | 0.024 |
| | Benzene | 1.005 | 0.16 | 0.240 | 0.989 | 0.019 | 0.060 |
| | | | | | | | |
| Systemic Lupus Erythematosus | PM10 | 1.031 | 0.136 | 0.240 | 0.996 | 0.880 | 0.912 |
| | PM2.5 | 1.042 | 0.131 | 0.240 | 0.997 | 0.930 | 0.930 |
| | NO2 | 1.011 | 0.005 | 0.015 | 1.008 | 0.395 | 0.474 |
| | NOx | 1.021 | 0.004 | 0.013 | 1.004 | 0.462 | 0.533 |
| | SO2 | 1.139 | 0.002 | 0.008 | 1.097 | 0.067 | 0.144 |
| | Benzene | 1.042 | 0.001 | 0.004 | 1.020 | 0.218 | 0.284 |
| | | | | | | | |
| Ankylosing Spondylitis | PM10 | 0.995 | 0.765 | 0.765 | 0.955 | 0.025 | 0.063 |
| | PM2.5 | 0.987 | 0.553 | 0.593 | 0.930 | 0.009 | 0.045 |
| | NO2 | 1.003 | 0.287 | 0.344 | 0.987 | 0.100 | 0.167 |
| | NOx | 1.007 | 0.312 | 0.360 | 0.993 | 0.095 | 0.167 |
| | SO2 | 1.142 | 0.000 | 0.000 | 1.042 | 0.316 | 0.395 |
| | Benzene | 1.013 | 0.253 | 0.316 | 0.980 | 0.164 | 0.246 |
| | | | | | | | |
| | PM10 | 0.944 | 0.000 | 0.000 | 0.913 | 0.000 | 0.000 |
| | PM2.5 | 0.924 | 0.000 | 0.000 | 0.879 | 0.000 | 0.000 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Psoriatic Arthritis | NO2 | 0.996 | 0.235 | 0.307 | 0.984 | 0.022 | 0.060 |
| | NOx | 0.993 | 0.198 | 0.270 | 0.991 | 0.018 | 0.060 |
| | SO2 | 1.044 | 0.154 | 0.240 | 0.992 | 0.818 | 0.909 |
| | Benzene | 0.991 | 0.381 | 0.423 | 0.977 | 0.08 | 0.150 |

[a] adjusted for age, gender, education, ethnicity, Townsend deprivation score, smoking and alcohol intake status and time spent outdoors and population density

### 5.6.1. Stratified analysis

Stratified analyses focused on the effects of $SO_2$ on vasculitis and rheumatoid arthritis with the aim of identify potential groups that may be at an increased risk. SO2 models were stratified by population density (urban vs rural), smoking (ever-smoked vs never-smoked), deprivation (least vs most deprived) and time spent outdoors in winter and summer (less than 3 hours vs more than 3 hours). The stratified models are reported for each select outcome, SV, and RA (Figure 5.4 and 5.5). In summary, the analyses based on population density (urban/rurality) showed that the effect of $SO_2$ on SV were higher in rural population (odds ratio: 1.179, 95% CI: 1.014 – 1.372), compared with population from urban areas (odds ratio: 1.047, 95% CI: 0.982 – 1.116). The effects of $SO_2$ on RA were also higher for those from rural areas (odds ratio:1.083, 95% CI: 1.005 – 1.167) compared with urban population (odds ratio: 1.029, 95% CI:1.001 – 1.058). Similarly, the effects of $SO_2$ were also increased in vasculitis and rheumatoid arthritis cases who were current or previous smokers, and those who reported to spend the most amount of time outdoors (> 3hours spent outdoors in winter and summer) including those the least deprived areas for RA.

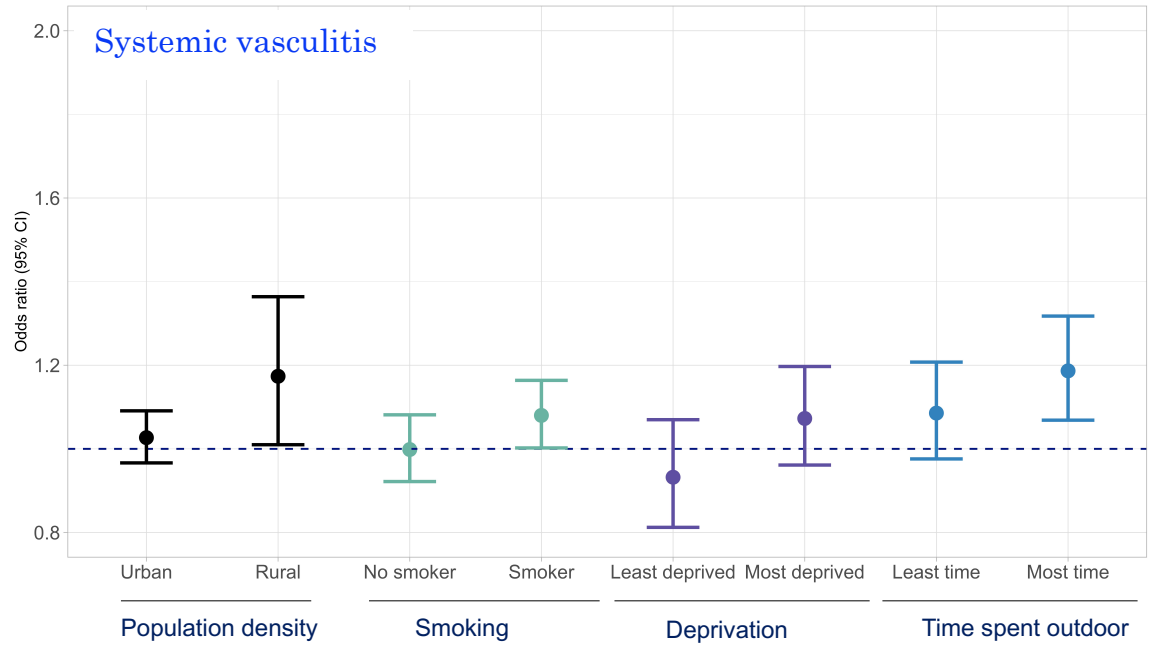



**Figure 5.4:** Effects of sulphur dioxide on systemic vasculitis stratified by population density, smoking, deprivation (Townsend score) and time spent outdoors

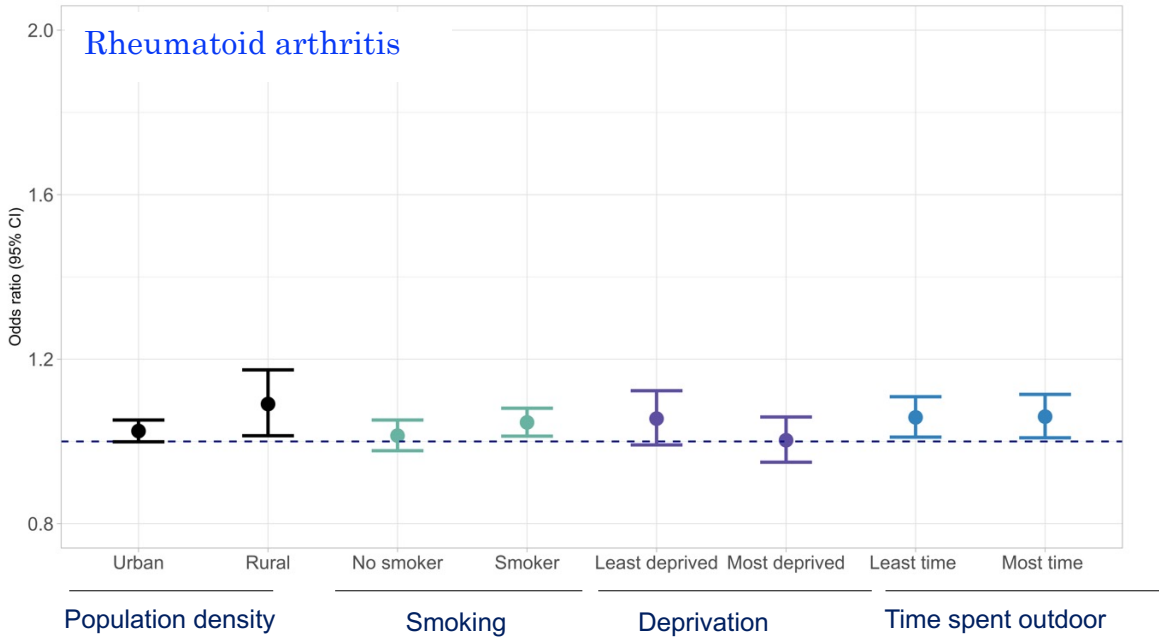



**Figure 5.5:** Effects of sulphur dioxide on rheumatoid arthritis risk stratified by population density, smoking, deprivation (Townsend score), and time spent outdoors.

### 5.6.2. Sensitivity analysis

Results from the sensitivity analyses were restricted to cases who had vasculitis or other autoimmune disease first occurring following entry into UK Biobank. A total of 802 participants with vasculitis, 3,588 participants with RA, 158 with SLE, 294 with ankylosing spondylitis and 571 psoriatic arthritis were identified (Table 5.3). Again, vasculitis cases were older (median age 61.7 years old) compared with the UK Biobank controls (median age = 56.46) and other autoimmune disease (combined median age = 57.84). 50% of vasculitis cases were smokers compared with 55% of UKB population. The time to event showed marginal variability between vasculitis cases and other autoimmune disease with a range of 4.73 years to 5.09 years. When the effects of air pollution were modelled using time to event analysis (cox regression), similar effects of $SO_2$ and pattern of association were observed for systemic vasculitis (adjusted hazard ratio: 1.070, 95%CI: 0.996 – 1.149) though this was not statistically significant due to the reduced sample size (Table 5.4). For other autoimmune diseases, $SO_2$ effects on rheumatoid arthritis (adjusted hazard ratio: 1.040, 95% CI: 1.005, 1.076) and systemic lupus erythematosus (adjusted hazard ratio: 1.117, 95% CI: 0.957,1.304) were similar to those observed in the main results (Figure 3), with statistical significance being seen for rheumatoid arthritis but not systemic lupus erythematosus.

**Table 5.3:** Characteristics of participants who developed systemic vasculitis and other autoimmune disease during follow up period in UK Biobank. These participants were included in the time to event (cox regression) results reported in Table 5.4.

| | **Systemic Vasculitis** ***n* =** 802 | **Rheumatoid Arthritis** ***n* =** 3,588 | **Systemic Lupus Erythematosus** ***n* =** 158 | **Ankylosing Spondylitis** ***n* =** 294 | **Psoriatic Arthritis** **n=**571 | **Population Controls** ***n* =** 487,914 |
|---|---|---|---|---|---|---|
| **Age** mean (±SD), years | 61.78 (6.06) | 59.27 (7.35) | 57.25 (7.93) | 58.18 (8.01) | 56.67 (7.76) | 56.46 (8.11) |
| **Gender** female, *n* (%) | 528 (65.84) | 2,433 (67.81) | 131 (82.91) | 142 (48.30) | 324 (56.74) | 223909 (45.9) |
| **Follow up (in years)** | | | | | | |
| Median (p25, p75) | 4.89 (2.951, 6.71) | 5.09 (3.052, 6.687) | 4.73 (2.997, 6.819) | 4.97 (2.609, 6.704) | 4.94 (2.343, 6.712) | 9.92 (9.232, 10.606) |
| **Smoking status** | | | | | | |
| Yes, n (%) | 401 (50.3) | 1,931 (54.3) | 84 (53.5) | 158 (54.5) | 310 (54.5) | 266,704 (55.0) |
| **Ethnicity,** *n* (%) | | | | | | |
| White | 759 (95.1) | 3337 (93.8) | 142 (90.4) | 277 (95.2) | 549 (96.8) | 459009 (94.6) |
| Asian or Asian British | 15 (1.9) | 91 (2.6) | 4 (2.5) | 3 (1.0) | 7 (1.2) | 9555 (2.0) |
| Black or Black British | 9 (1.1) | 73 (2.1) | 9 (5.7) | 4 (1.4) | 1 (0.2) | 7831 (1.6) |
| Chinese | 3 (0.4) | 10 (0.3) | 0 (0.0) | 0 (0.0) | 3 (0.5) | 1541 (0.3) |
| Mixed or other ethnic groups | 12 (1.5) | 47 (1.3) | 2 (1.3) | 7 (2.4) | 7 (1.2) | 7301 (1.5) |
| **Close to major road (2008)** | | | | | | |
| Median (p25, p75) | 45 (5.7) | 285 (8.1) | 13 (8.3) | 24 (8.3) | 40 (7.1) | 34303 (7.1) |

**Table 5.4**: Sensitivity analysis of time to event (Cox regression) restricted to participants who were diagnosed or had a vasculitis or other autoimmune disease occurring after entry into UKB.

| | Pollutant | Unadjusted | | | | Adjusted [b] | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | HR | 95% CI | | p-value | HR | 95% CI | | p-value |
| Systemic Vasculitis | PM10 | 0.956 | 0.929 | 0.984 | 0.002 | 0.969 | 0.934 | 1.005 | 0.093 |
| | PM2.5 | 0.939 | 0.904 | 0.976 | 0.001 | 0.955 | 0.909 | 1.004 | 0.069 |
| | NO2 | 0.991 | 0.980 | 1.002 | 0.104 | 0.996 | 0.982 | 1.010 | 0.543 |
| | NOx | 0.995 | 0.990 | 1.001 | 0.129 | 0.998 | 0.990 | 1.005 | 0.554 |
| | SO2 | 1.020 | 0.960 | 1.084 | 0.518 | 1.070 | 0.996 | 1.149 | 0.064 |
| | Benzene | 1.070 | 0.996 | 1.149 | 0.064 | 1.024 | 0.901 | 1.164 | 0.717 |
| Rheumatoid Arthritis | | | | | | | | | |
| | | HR | 95% CI | | p-value | HR | 95% CI | | p-value |
| | PM10 | 0.985 | 0.972 | 0.998 | 0.027 | 0.965 | 0.949 | 0.981 | 0.000 |
| | PM2.5 | 0.975 | 0.958 | 0.993 | 0.006 | 0.946 | 0.925 | 0.968 | 0.000 |
| | NO2 | 1.003 | 0.998 | 1.008 | 0.281 | 0.994 | 0.987 | 1.000 | 0.055 |
| | NOx | 1.001 | 0.999 | 1.004 | 0.378 | 0.997 | 0.993 | 1.000 | 0.045 |
| | SO2 | 1.087 | 1.057 | 1.117 | 0.000 | 1.040 | 1.005 | 1.076 | 0.023 |
| | Benzene | 1.017 | 0.970 | 1.066 | 0.487 | 0.938 | 0.884 | 0.996 | 0.036 |
| | | | | | | | | | |
| Systemic Lupus Erythematosus | | HR | 95% CI | | p-value | HR | 95% CI | | p-value |
| | PM10 | 1.037 | 0.975 | 1.103 | 0.244 | 0.999 | 0.925 | 1.079 | 0.976 |
| | PM2.5 | 1.047 | 0.964 | 1.137 | 0.278 | 0.995 | 0.897 | 1.104 | 0.926 |
| | NO2 | 1.027 | 1.004 | 1.051 | 0.020 | 1.009 | 0.979 | 1.039 | 0.565 |
| | NOx | 1.013 | 1.001 | 1.025 | 0.029 | 1.003 | 0.988 | 1.019 | 0.663 |
| | SO2 | 1.189 | 1.049 | 1.348 | 0.007 | 1.117 | 0.957 | 1.304 | 0.160 |
| | Benzene | 1.288 | 1.063 | 1.561 | 0.010 | 1.131 | 0.879 | 1.456 | 0.339 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Ankylosing Spondylitis | | **HR** | **95% CI** | | **p-value** | **HR** | **95% CI** | | **p-value** |
| | PM10 | 0.993 | 0.948 | 1.040 | 0.753 | 0.954 | 0.899 | 1.012 | 0.118 |
| | PM2.5 | 0.987 | 0.928 | 1.050 | 0.689 | 0.936 | 0.864 | 1.014 | 0.105 |
| | NO2 | 1.008 | 0.990 | 1.026 | 0.370 | 0.985 | 0.963 | 1.008 | 0.208 |
| | NOx | 1.004 | 0.995 | 1.013 | 0.404 | 0.992 | 0.980 | 1.004 | 0.191 |
| | SO2 | 1.151 | 1.047 | 1.265 | 0.003 | 1.038 | 0.921 | 1.170 | 0.537 |
| | Benzene | 1.130 | 0.968 | 1.319 | 0.123 | 0.978 | 0.799 | 1.198 | 0.833 |
| | | | | | | | | | |
| Psoriatic Arthritis | | **HR** | **95% CI** | | **p-value** | **HR** | **95% CI** | | **p-value** |
| | PM10 | 0.951 | 0.919 | 0.984 | 0.004 | 0.929 | 0.890 | 0.970 | 0.001 |
| | PM2.5 | 0.931 | 0.890 | 0.974 | 0.002 | 0.899 | 0.848 | 0.952 | 0.000 |
| | NO2 | 0.995 | 0.982 | 1.008 | 0.432 | 0.988 | 0.972 | 1.005 | 0.168 |
| | NOx | 0.997 | 0.990 | 1.004 | 0.399 | 0.994 | 0.985 | 1.002 | 0.152 |
| | SO2 | 1.042 | 0.971 | 1.119 | 0.255 | 1.001 | 0.918 | 1.090 | 0.991 |
| | Benzene | 0.964 | 0.854 | 1.089 | 0.558 | 0.926 | 0.796 | 1.077 | 0.316 |

[b] age, gender, education, ethnicity, Townsend deprivation score, smoking and alcohol intake status and time spent outdoors and population density

### 5.6.3. Association between air pollution and markers of inflammation and disease activity at baseline

As mentioned in the overview section of this chapter, this thesis took advantage of UKB wide measurement of biomarkers of various diseases and outcomes. For this chapter, we assessed the short-term impact of air pollution (annual mean exposure) on the expression and frequency of blood markers of inflammation and disease activity. These analyses will support the growing hypothesis that air pollution is an inflammatory trigger which can cumulative systemic effects that can lead to inflammatory autoimmune diseases. Results of these analyses are summarised in Figure 3 using a feature expression heat map, a visual method used to explore complex associations between multiple variable sets. In summary, there was a statistically significant association between annual sulphur dioxide with log concentrations of CRP, neutrophil, and monocyte. A 1 µg per cubic meter increase in sulphur dioxide was associated with 3.2% (95% CI: 2.93 – 3.44) lower concentration of CRP and 0.8% (95% CI: 0.79 – 0.96) and 0.8% (95% CI: 0.77 – 0.94) higher in total neutrophil and monocyte count. Conversely, benzene, $PM_{10}$ and $PM_{2.5}$ were inversely association with CRP and eosinophil while no association were observed for the nitrogen oxides ($NO_2$ and $NO_x$) pollutants

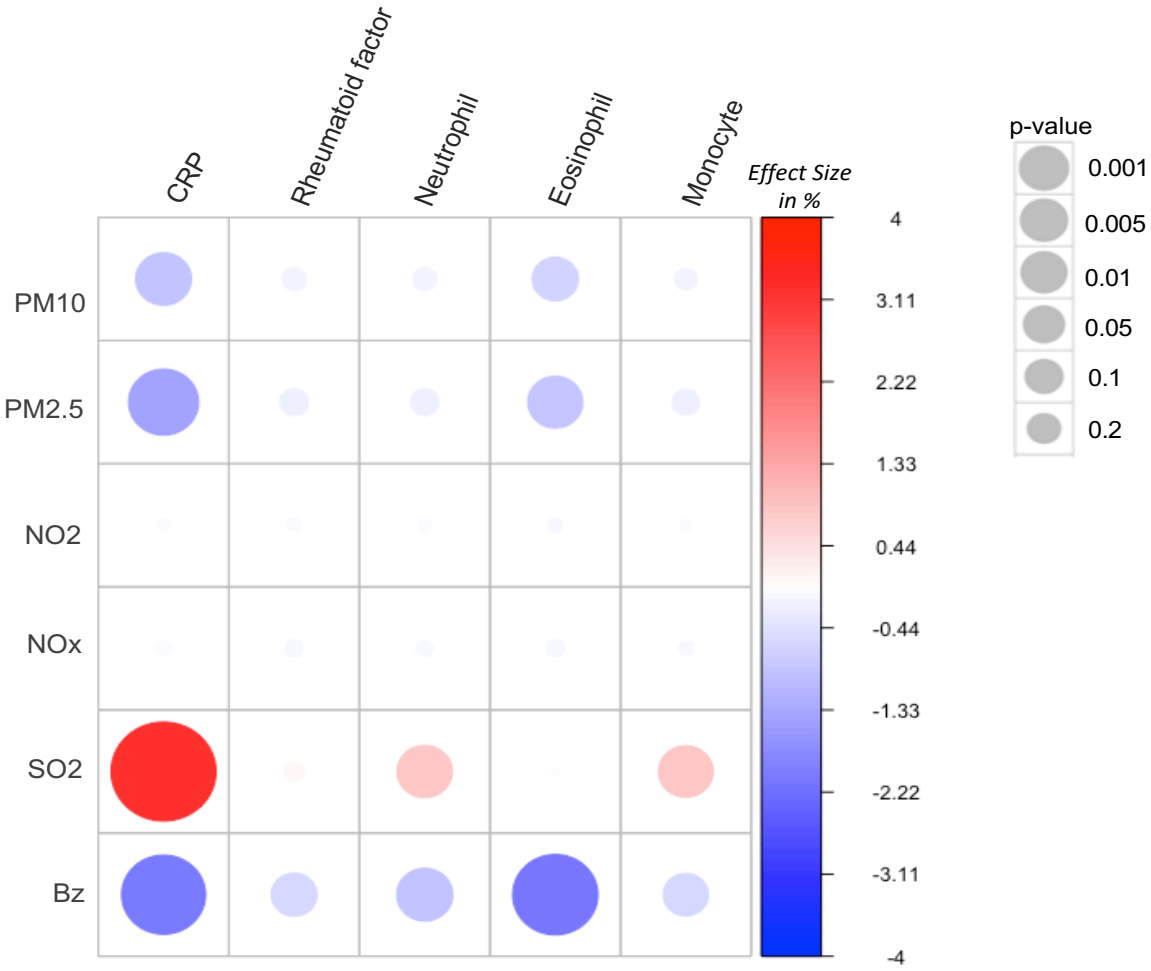



**Figure 5.6:** A feature expression heatmap highlighting the relationship between air pollution (annual mean) exposure at baseline (2006-2010) and overall percentage increase/decrease in markers of inflammation in UK Biobank. The effect sizes are reported in percentages and are calculated by multiplying the standardised (beta) coefficients from a linear regression model with 100.

## 5.7. Discussion

This is the first large population study to quantify the long-term impact of outdoor air pollution on vasculitis. It is also the first to provide a robust analyses that compares of the effect of air pollution across multiple inflammatory autoimmune diseases. The findings of this chapter show that vasculitis patients in UK Biobank were more likely be female, to be without a college education and from more deprived areas compared with the rest of the UK Biobank cohort. This demographic disparity in vasculitis population based on measures socioeconomic status have not been describe in detail before. There is only one study that has attempted to assess this (Naz et al.,2019) in a small study population from Cheshire, UK, and reported no clear association between socio-economic measures of employment and the prevalence of vasculitis (97).

Furthermore, the standardised incidence of vasculitis reported here were lower than expected, this is compared with previous estimates from previous studies in the UK(95,96,310–312). This difference in incidence estimates may reflect variation in methods used to define vasculitis. For instance, in the study reported here, vasculitis was defined based on a broad clinical categorisation using the three-character ICD codes that pooled all small to large vasculitis into one disease entity.

Most importantly, a major findings of this study shows that sulphur dioxide is significantly associated with vasculitis. This is the first study to show this association and to provide a comparison with other autoimmune disease that indicates that the effects of $SO_2$ are consistent across major inflammatory autoimmune disease, like rheumatoid arthritis, and systemic lupus erythematosus. Furthermore, the effects of $SO_2$ were again seen with markers of inflammation, particularly CRP, neutrophil and monocyte count. At greater risk were individuals from rural areas, smokers, and those who reported to spend more than 3 hours outdoor both in winter and summer. Additionally, $PM_{10}$ and $PM_{2.5}$ showed inverse associated with rheumatoid arthritis, ankylosing

spondylitis and psoriatic arthritis. While no clear associations were seen for all the other pollutants ($NO_2$, NOx and Benzene) assessed in this study.

Compared to previous studies, the findings of $PM_{10}$ and $PM_{2.5}$ show a similar pattern with previous studies reporting effects of air pollution on rheumatoid arthritis. For instance several studies have reported that an inverse pattern association between particulate matter and RA, even though most of these studies were not statistically significant due to limited power and sample size (313,314). As for other autoimmune disease outcomes, there are no studies reporting effect of air pollution on ankylosing spondylitis and psoriatic arthritis. The inverse effects seen with particulate matter are first and cannot be compared with other studies to date.

Importantly, the finding of sulphur dioxide and its association with vasculitis and rheumatoid arthritis comes amid growing evidence that report the role of outdoor air pollution as potential risk factor of autoimmune diseases (315), with another body of literature indicating that certain air pollutants may important predictor of annual changes in biomarkers of inflammation and immune response(314,316,317). For instance, a recent study by Berlatsky et al.,2020 showed that $SO_2$ had a small but significant association with expression anti-citrullinated antibodies (ACPA) (Odds ratio: 1.02, 95%CI:1.00-1.04) in large cohort of people recruited in the CARTaGENE biobank in Canada, (318,319). These findings were further dissected in a separate study using the same cohort to specifically assess the impact of the environment on the whole transcriptomic gene expression and health related traits by using air pollution as lead measure of environment (320). Overall, the major finding of this study showed that $SO_2$ was significantly associated with 170 expressed genes involved in leukocyte migration, CXCR chemokine activity and G-protein-coupled receptors during chronic inflammation. The study also showed that for individuals residing in areas with high industrial activity, $SO_2$ was associated with increased blood concentration of gamma-glutamyl transferase (GGT), and monocyte count.

Mechanistic studies on gene-environment interaction, have highlighted the role major air pollutants, like $PM_{10}$ and $PM_{2.5}$, as potential mediator of oxidative stress and DNA damage by inducing alterations in membrane channel at the site of initial exposure, specifically the lungs and small to medium blood vessels (321,322). Animal studies have also shown a dose-gradient response in TNF-alpha, IL1-beta expression following a 7 day repeat exposure to $SO_2$ on in the lungs and heart(323). Similar association have also been reported in human controlled exposure studies where controlled levels of $SO_2$ were exposed to study participants, leading to a lagged increases in IL-6 concentration in the blood(324) and neutrophil recruitment and activation in lung tissue and as well as peripheral in other studies(325).

A major strength of the present study is its power and large sample size and geographic coverage that reflect urban and rural geographies of UK. The study also considered a wide range of potential confounders, each measured in a standard way, with a same list of confounders being adjusted for across multiple autoimmune disease reported here. The reported effect was rigorously evaluated in sensitivity analyses under different statistical method and assumptions. Additionally, the exposure measures used here corresponded to each participants place of residence at between 2001 to 2018 and captured any movement to new address.

As outlined in previous chapters, several limitation must be considered when interpreting the result of UK Biobank. First, the outcome data used here were limited to three-character ICD codes. We did not have access to the four-character ICD codes that would allow for identification of the different vasculitis clinical phenotypes. Systemic vasculitis as an outcomes was derived by pooling all small to large vessel vasculitis as one disease entity using ICD M30-31 code. It is not possible to infer whether the effects seen for $SO_2$ would be true for different vasculitis phenotypes as well other autoimmune diseases subtypes. Furthermore, it is possible that vasculitis cases diagnosed between 2001 and 2005 had a higher air pollution exposure than the rest of the cohort, possibly introducing a temporal bias. The pollution exposure used here refer to outdoor exposure and may not indicate true exposure as most people majority of their

time indoors. Lastly, cases derived from the first occurrences data fields, sourced self-reported health conditions recorded at baseline are subject to recall bias. Importantly, UK Biobank is subject to ‘representative’ bias as it consists of participants who were mainly females, to live in less socioeconomically deprived areas and had better health seeking behaviour compared with the UK general population (326). UKB may therefore not be generalisable to the rest of the UK population.

## 5.8. Conclusion and next steps

This present study reported first novel results showing a significant association between sulphur dioxide and an increased risk of vasculitis and rheumatoid arthritis. Additional analyses on air pollution and biomarkers of inflammation provide indication that sulphur dioxide may serve as an inflammatorily trigger with cumulative exposure. Given the limitation highlighted in the discussion regarding how vasculitis was defined in UK Biobank, the next chapter will seek to validate the results of the present chapter by using routine data from the Scottish Morbidity of cases diagnosed with ANCA-associated vasculitis and Giant cell arteritis with goal to understand if the specific effect of air pollution on vasculitis.

# Chapter 6- Geographic variation of AAV occurrences in Scotland and its association with residential air pollution – A validation study

## 6.1.Overview

The objective of this chapter is to provide a validation assessment of the effect of air pollution on vasculitis, with a specific focus on ANCA-Associated vasculitis and Giant cell arteritis. Chapter 5 revealed from UK Biobank an important association between sulphur dioxide and the onset of vasculitis. The results from the previous chapter were limited the way vasculitis was classified. Vasculitis cases were classified based on three-character ICD coding (ICDM30 and ICDM31). These codes pooled all vasculitis phenotype into one a single outcome. To address these limitations, this study used hospital admission data from the Scottish morbidity record (SMR01) and focused on AAV and GCA. Air pollution data from DEFRA was linked to the SMR01 with the aim of assessing the association between air pollution and the onset of AAV and GCA. Stratified analyses were carried based on the Scottish urban or rural geographies with the aim of investigating the geographic variation of air pollution effects on vasculitis. This was done in light with growing evidence to indicate that the prevalence of vasculitis may vary by urban and rural geographies from Chapter 1. Aiyegbusi et al., 2021 recently showed that the incidence of biopsy-proven GPA was associated with rurality in Scotland (98). Similarly, Anderson et al., 2013 reported higher incidence of AAV in the northern rural part of Saskatchewan, Canada(171). Ormerod et al., 2008 also showed an increase in frequency of MPA in rural Australia(327). There is also some evidence to suggest that populations living near quarries or an open pit mine may be at a greater risk for AAV. Giorguitti et al.,2021 recently showed that reported in residents living near a quarry in the northeast region of France had 2.5-fold increased odds of developing GPA including PR3+ and MPO+ AAV (172). It is not clear whether these geographic differences in AAV occurrence are attributable to air pollution. Lessons from the literature on other vasculitis show spatiotemporal variation in clusters of Kawasaki disease (KD) may be explained

by environmental exposures. One particular example is the evidence from studies in the US and Japan (328–330). Specifically, Sano et al.,2016 showed that the geographic differences in clusters of Kawasaki disease seen between 2007 and 2012 in Tokyo and Kumamoto, Japan could be attributed to airborne exposures (328). For the study in the US, Rypdal et al.,2018 hypothesised that the spatiotemporal cluster seen in KD in California based on time scales of up to 10 days and spatial scales of 10 to 100 km was partly modulated by regional and historic differences in weather conditions (331). Besides the number of studies reporting the effect of geography on AAV and Kawasaki disease, there are no studies assessing this in GCA. This chapter will provide the first major analyses showing the geographic effect of air pollution on AAV and GCA while adjusting for limited number of potential confounders available in routine health dataset in Scotland.

## 6.2.Method

### 6.2.1. Study population

The data used in this chapter were sourced from the Information Services Division (ISD), a division of the National Services Scotland (NSS), part of the National Health Services (NHS) and Public Health Scotland (PHS). The ISD hosts the National Data Catalogue (NDC) which serves as the single definitive resource of information on Scottish health and social care. Healthcare interactions at primary, secondary and tertiary care are comprehensively computerised and are derived 14 NHS heath boards covering the Scottish population. In Scotland, secondary and tertiary care are delivered mostly in hospital settings, as inpatient, outpatient and day case admissions or episodes. This data is held by ISD as Scottish morbidity records (SMR01) and are recorded as hospital episodes with date admission and discharge. For this study, patients with ANCA-Associated vasculitis, and giant cell arteritis were the focus of the analyses reported here. All other vasculitis were pooled into one disease category due to their small number of cases which may be potentially identifiable and therefore would not be disclosed by ISD. Estimated prevalence

of systemic vasculitis ranges from 0.1 per 100 000 to 2.6 per 100 000 population for ANCA-associated vasculitis (AAV) and 25 per 100,000 (Giant Cell Arteritis) and to date more than 1,700 AAV and 4000 GCA patients have been identified by the ISD team. All patients with AAV and GCA in Scotland were identified from the national administrative databases (SMR01)using the following ICD-10 codes (below), used in Scotland since April 1996, which provides greater certainty and specificity than ICD-9 codes:

ANCA-Associated vasculitis

(a) Granulomatosis with Polyangiitis (GPA - M31.3)

(b) Microscopic Polyangiitis (MPA - M31.7)

(c) Eosinophilic granulomatosis with polyangiitis [Churg-Strauss]/EGPA - M30.1)

Giant cell arteritis

(d) Giant Cell Arteritis (GCA - M31.5 and M31.6)

Other vasculitis

(e) Polyarteritis nodosa (M30.0)

(f) Other conditions related to polyarteritis nodosa (M30.8)

(g) Arteritis, unspecified (I77.6)

(h) Other specified disorders of arteries and arterioles (I77.8)

It is worth noting, due to limited specificity of the ICD coding, the other vasculitis group, particularly PAN may include MPA patients because of the changing classification and nomenclature system. A control group representative of the general population was sourced from the CHI register and SMR01. Since 1970, each person registered with a GP in Scotland has been assigned a unique identification number, known as the Community Health Index (CHI) which is held under the national CHI population register(332). CHI register holds information about people's name, postcode, general practice, date and region of registration and date of death, where applicable. Although generated for the purpose of primary care, the CHI number is used in tertiary care system to provide follow up data on a person's trajectories throughout the healthcare system. The control group was identified by the ISD team and consisted of more than 120,000 individuals from the CHI register and SMR01. For this chapter and analyses, the control was unmatched by age or sex and consisted of a

random sample of individual over the age 16 years old in Scotland. Access to the Scottish vasculitis cohort was achieved under the VOICES study which received ethics approval from the Scottish national NHS ethics committee and approvals from the Public Benefit and Privacy panel for Health and Social care (PBPP 1819-0069 – Appendix 4 and 5)

### 6.2.2. Data linkage

Two type of data linkage were carried, one, for electronic administrative health record and the second, for the environmental data. The linking of administrative health records held by the ISD was facilitated by the electronic Data Research and Innovation Service (eDRIS) team. In line with our research goals, the eDRIS provided the linkage of the following datasets consisting of important demographic measures that included confounding measures which need to be considered when adjusting for the impact of environmental factors on vasculitis:

(a) Scottish Morbidity Record (SMR) – records of all hospital inpatient and day case admissions, (SMR01), enabled the identification of cases, controls, and included details on mortality.
(b) CHI register– enabled matching of population control group and provided confirmation of GP registration
(c) NRS Death Register (held in eDRIS) – mortality register
(d) Scottish Index of Multiple Deprivation – Deprivation Measures
(e) Scottish Government Urban Rural Classification – Population Density

The de-identified linked data were analysed in the National Safe haven and is retained by eDRIS for the duration of study period. To enable assessment of linkage and data extraction quality, validation metrics including duplicate extraction of small number of variables across datasets (e.g.: gender) and expected pairings (disease plus prescription pairing) were used as positive and negative controls for the linkage process. This was conducted centrally by the eDRIS linkage team. The environmental linkage was achieved by transforming the patient's postcode sector (4 to 5 character postcode) into grid coordinates (easting and northing) and linking it with annual pollution measures, which were

modelled estimates derived from the Pollution Climate Mapping models developed by UK department for Environment, Food and Rural Affairs (DEFRA) (https://uk-air.defra.gov.uk/data/pcm-data). Chapter 4 gave a summary of the methods that can the linking of environmental data. Air pollution measure were estimated on a 1kmx1km square grid which has an associated central point (centroid). The pollution concentration provided were calculate at an intermediate zone level around the centroid. A mean estimate for that zone is then given and patient within that zone were assigned a measure of exposure based on the year of diagnosis (2001 to 2020). For zones where there were no grid square centroids, the pollution estimates from the nearest grid square were used.

### 6.2.3. Statistical analysis

All the analyses were performed using R studio v1.2.5019 and Stata 13. Baseline characteristic for participants with systemic vasculitis and other autoimmune diseases are summarised using descriptive statistics. Quintiles were generated for Scottish index of multiple deprivation. For normally distributed variables (age), mean and standard deviation is reported. For variables with skewed distribution (pollution), median and interquartile range were calculated and is reported. Relative frequencies in percentage are reported for categorical data. Logistic regression was used to assess the association between air pollution exposures and ANCA-Associated vasculitis, giant cell arteritis, other vasculitis as well as population controls. All the outcomes reported were modelled (1-cases vs 0-population control) separately, first unadjusted and adjusted for potential confounders (age, sex, Scottish index of multiple deprivation and population density). All the covariables included in the modelling stage were within reasonable range of multicollinearity (variation inflation factor of 1.2 to 1.7) and were selected to minimise confounding. Given there were a total 18 models for the outcome of interest, the two-tail p-values were adjusted using Simes-Benjamini-Hochberg false discovery rate (FDR at 0.05 significance level) method, both for the unadjusted and adjusted. Additionally, a stratified analysis was undertaken to get insight into geographic

variation of air pollution by Scottish health board of residence. This consisted of fitting separate models (unadjusted and adjusted for potential confounders) stratified by urban and rural classification for AAV,GCA and other vasculitis.

## 6.3.Results

### 6.3.1. Study characteristics

**Table 7.1** provides a summary of the characteristics of Scottish vasculitis population compared with controls from the general population. In summary, a total of 1,703 of AAV, 4,125 GCA, 2,094 other vasculitis patients and 120,962 population controls were identified from the VOICES study. GCA patients were older (mean age, 72,65, sd±9.74) and with a high proportion being female (70.9%) compared with AAV (mean age, 58.74 , 49.9% female) and other vasculitis patients (mean age 62.50, and 55.2% female), respectively. AAV and GCA patients showed on average to be from most deprived areas (21.7 and 20.3%) compared with other vasculitis patients (17.8%). Overall, most of the vasculitis and population controls were from urban areas Most of the cases were from the largest health boards in Scotland - NHS Greater Glasgow and Clyde (health board 7), NHS Lothian (health board 6), and NHS Grampian (Board 8), which accounted for more than 45% of AAV cases, 50% of GCA cases and 38% of other vasculitis cases. Smaller health boards like NHS Orkney (health board 11) , Shetland (health board 12) and Western Isles (health board 14) accounted for less 1% of AAV, GCA and other vasculitis each. Exploratory analyses assessing the distribution of air pollution by deprivation highlighted potential demographic disparities in air pollution between AAV and GCA compared with population controls. Figure 6.2 and 6.3 shows a summary of these analyses. Overall, air pollution exposure was highest among the two extremes of deprivation status group (least and most deprived) regardless of disease status. GCA and other vasculitis cases from the most and least deprived areas, showed a higher $PM_{10}$ and $PM_{2.5}$ exposures compared with population controls. Interestingly, AAV patients from the most and least deprived area showed higher $SO_2$ exposure compared with population

controls, while GCA patients had higher benzene exposures compared with population controls.

**Table 6.1:** Cohort characteristics of patients with ANCA-vasculitis, Giant cell arteritis, other vasculitis, and the Scottish population controls in the VOICES study

| | **ANCA-Associated Vasculitis [a]** ***n* = 1,703** | **Giant Cell Arteritis** ***n* = 4,125** | **Other Vasculitis [b]** ***n* = 2,094** | **Scottish population Controls** ***n* = 120,962** |
|---|---|---|---|---|
| **Age** mean (±SD), years | 58.74 (±15.68) | 72.65 (±9.74) | 62.50 (±16.17) | 69.17 (±14.06) |
| **Gender** female, *n* (%) | 850(49.9) | 2,926 (70.9) | 1,155 (55.2) | 73,805 (61.0) |
| **Index Multiple deprivation** | | | | |
| 1 Most deprived, *n* (%) | 299 (17.6) | 766 (18.6) | 486 (23.2) | 25231 (20.9) |
| 5 Least deprived | 369 (21.7) | 839 (20.3) | 320 (15.3) | 21483 (17.8) |
| **Population density** | | | | |
| Urban (%) | 1,315 (77.22) | 3301 (80.12) | 1642 (78.41) | 95,271 (78.76) |
| Rural | 388 (22.78) | 824 (19.98) | 452 (21.59) | 25,691 (21.24) |
| **Health board of residence, *n(%)*** | | | | |
| 1 | 144(8.5) | 275 (6.6) | 160 (7.6) | 8,516 (7.0) |
| 2 | 36(2.1) | 71 (1.7) | 55 (2.6) | 3,010 (2.5) |
| 3 | 59(3.5) | 153 (3.7) | 69 (3.3) | 4,727 (3.9) |
| 4 | 135(7.9) | 238 (5.8) | 155 (7.4) | 8,374 (6.9) |
| 5 | 86 (5.0) | 200 (4.8) | 106 (5.1) | 5,797 (4.8) |
| 6 | 211 (12.4) | 375 (9.1) | 165 (7.9) | 11,600 (9.6) |
| 7 | 312 (18.3) | 783 (19.0) | 401 (19.1) | 23,282 (19.2) |
| 8 | 134 (7.9) | 265 (6.4) | 152 (7.2) | 10,089 (8.3) |
| 9 | 160 (9.3) | 347 (8.4) | 286 (13.7) | 11,063 (9.1) |

| | | | | | |
|---|---|---|---|---|---|
| | 10 | 253 (14.9) | 875 (21.2) | 223 (10.6) | 21,187 (17.5) |
| | 11 | <20 (<1) | <20 (<1) | <20 (<1) | 548 (<1) |
| | 12 | <20(<1) | 18 (<1) | <20 (<1) | 713 (<1) |
| | 13 | 144 (8.5) | 493 (12.0) | 292 (13.9) | 11,085 (9.2) |
| | 14 | <20 (<1) | <20(<1) | <20 (<1) | 971 (<1) |
| **PM10 ($_{2001-2020}$) (µg/m3)** | | | | | |
| Median | | 11.03 | 11.23 | 11.07 | 11.07 |
| (p25; p75) | | (9.87,11.95) | (10.08, 12.11) | (9.92, 12.00) | (9.87, 12.00) |
| **PM2.5 $_{(2002-2020)}$ (µg/m$^3$)** | | | | | |
| Median | | 6.96 | 7.04 | 6.98 | 6.98 |
| (p25; p75) | | (6.15, 7.94) | (6.34, 7.64) | (6.25, 7.51) | (6.18, 7.54) |
| **NO2 $_{(2002-2020)}$ (µg/m$^3$)** | | | | | |
| Median | | 11.38 | 12.10 | 11.50 | 11.38 |
| (p25; p75) | | (6.59, 15.33) | (7.37, 16.27) | (6.58, 15.82) | (6.56, 15.91) |
| **NOx $_{(2002-2020)}$ (µg/m$^3$)** | | | | | |
| Median | | 15.68 | 16.82 | 15.91 | 15.68 |
| (p25; p75) | | (8.82, 22.69) | (9.93, 24.10) | (8.80, 23.35) | (8.80, 23.36) |
| **SO2 $_{(2002-2020)}$ (µg/m$^3$)** | | | | | |
| Median | | 1.62 | 1.64 | 1.63 | 1.63 |
| (p25; p75) | | (1.22, 2.08) | (1.36, 2.03) | (1.31, 2.11) | (1.26, 2.02) |
| **Benzene $_{(2003-2020)}$ (µg/m$^3$)** | | | | | |
| Median | | 0.32 | 0.32 | 0.31 | 0.32 |
| (p25; p75) | | (0.23, 0.40) | (0.24, 0.40) | (0.23, 0.39) | (0.23, 0.40) |

[a.] includes GPA, MPA, EGPA, [b.] includes PAN, Other conditions related to PAN, Arteritis, unspecified, Other specified disorders of arteries and arterioles

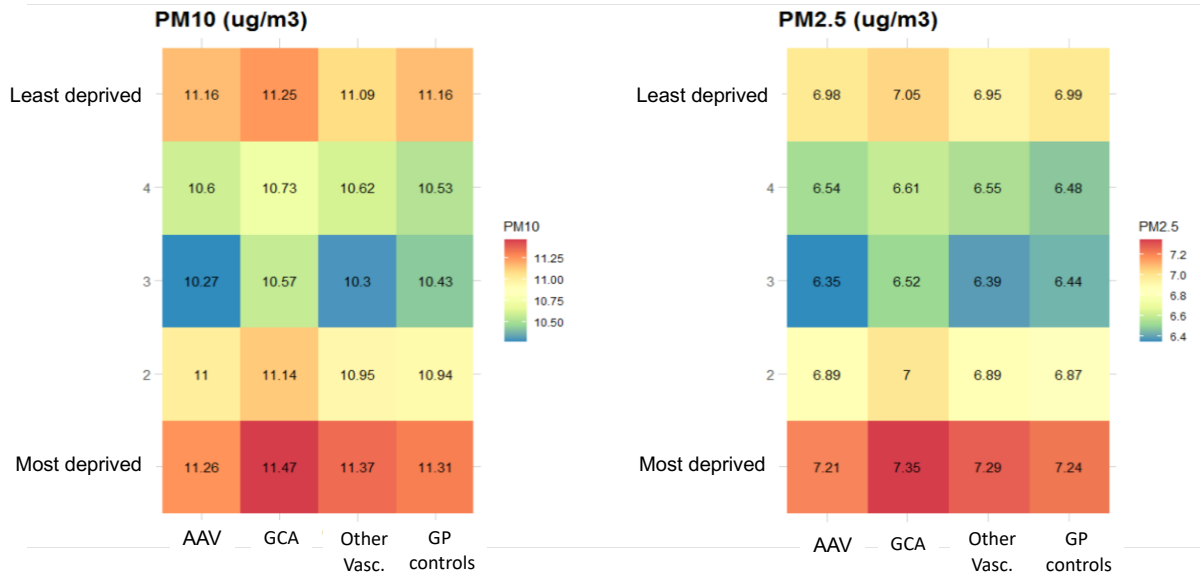



**Figure 6.1:** Heatmap based on the Scottish deprivation index shows the median exposure of particulate matters (PM10 and PM2.5) from 2001 to 2020 in people with AAV, GCA, and other vasculitis compared with general population controls.

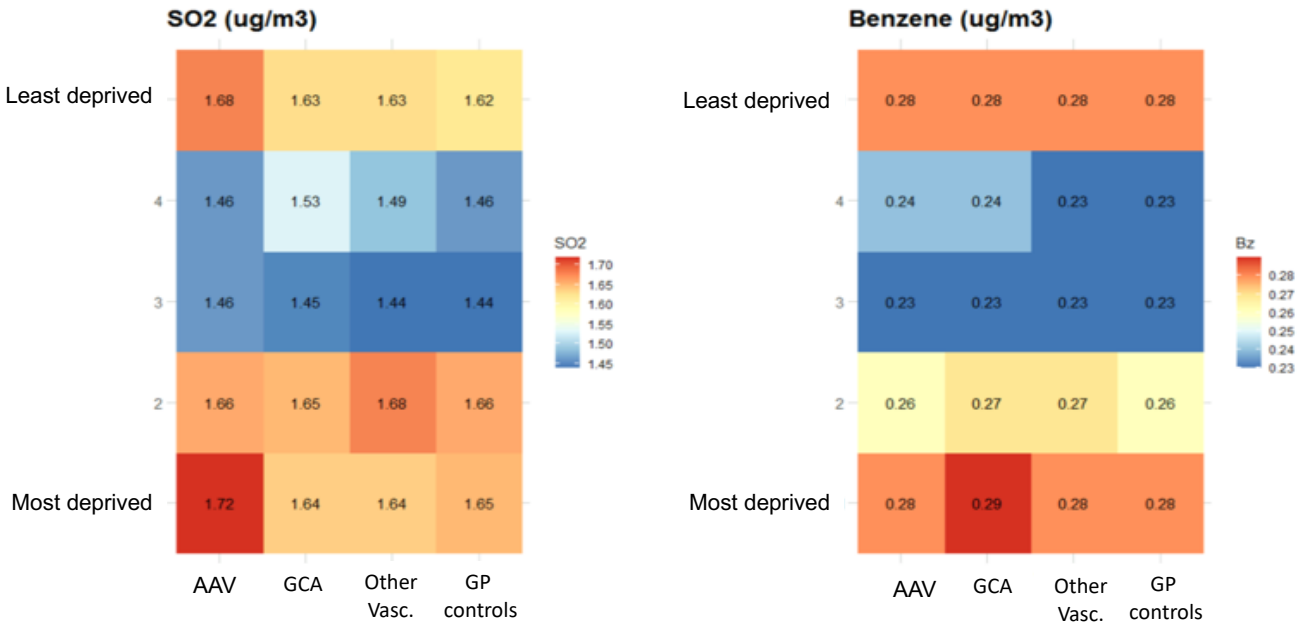



**Figure 6.2:** Heatmap based on the Scottish deprivation index shows the median sulphur dioxide ($SO_2$) and benzene ($C_6H_6$) exposure between 2001 to 2020 in people with AAV, GCA, and other vasculitis compared with the population controls from the SMR01.

### 6.3.2. Association between air pollution and vasculitis risk in Scotland

Analyses quantifying the relationship between air pollution exposures and vasculitis shows sulphur dioxide and particulate matter are important pollutants particularly associated with AAV and GCA. Figure 6.2 and Table 6.1 provides a summary of fifteen models, adjusted for age, sex, index multiple deprivation, and population density, depicting this association between long-term air pollution exposure and a risk for ANCA-Associated, Giant cell arteritis and other systemic vasculitis. Specifically, sulphur dioxide showed significant association with AAV and Other vasculitis while particulate matter ($PM_{10}$ and $PM_{2.5}$) was associated with GCA. Additionally, benzene was associated with AAV and GCA. For AAV and other vasculitis, 1 µg per cubic meter higher average level of $SO_2$ in the period 2002-2020 was associated with 6.86% (adjusted odds ratio: 1.0686, 95% CI: 1.007 – 1.134) and 7.6% (adjusted odds ratio: 1.076, 95% CI: 1.019 - 1.136) increased odds of other vasculitis. For GCA, 1 µg per cubic meter higher average level of in $PM_{10}$ and $PM_{2.5}$ was associated with 6.94% (adjusted odds ratio: 1.069, 95% CI: 1.044 – 1.095) and 11% (adjusted odds ratio: 1.107, 95% CI: 1.068 – 1.148) increased odds in disease onset. Similarly for benzene, 1 µg per cubic meter higher average level of was associated with 3.2% (adjusted odds ratio: 1.0323, 95% CI: 1.002 – 1.064) and 2.4% (adjusted odds ratio: 1.024, 95% CI: 1.004 -1.047) increase odds of AAV and GCA, respectively. Nitrogen oxides ($NO_2$ and $NO_x$) had a statistically significant increased risk effect of GCA that ranged from 0.6 to 1%. For AAV and Other vasculitis, there were not statistically significant association with nitrogen dioxide or oxide.

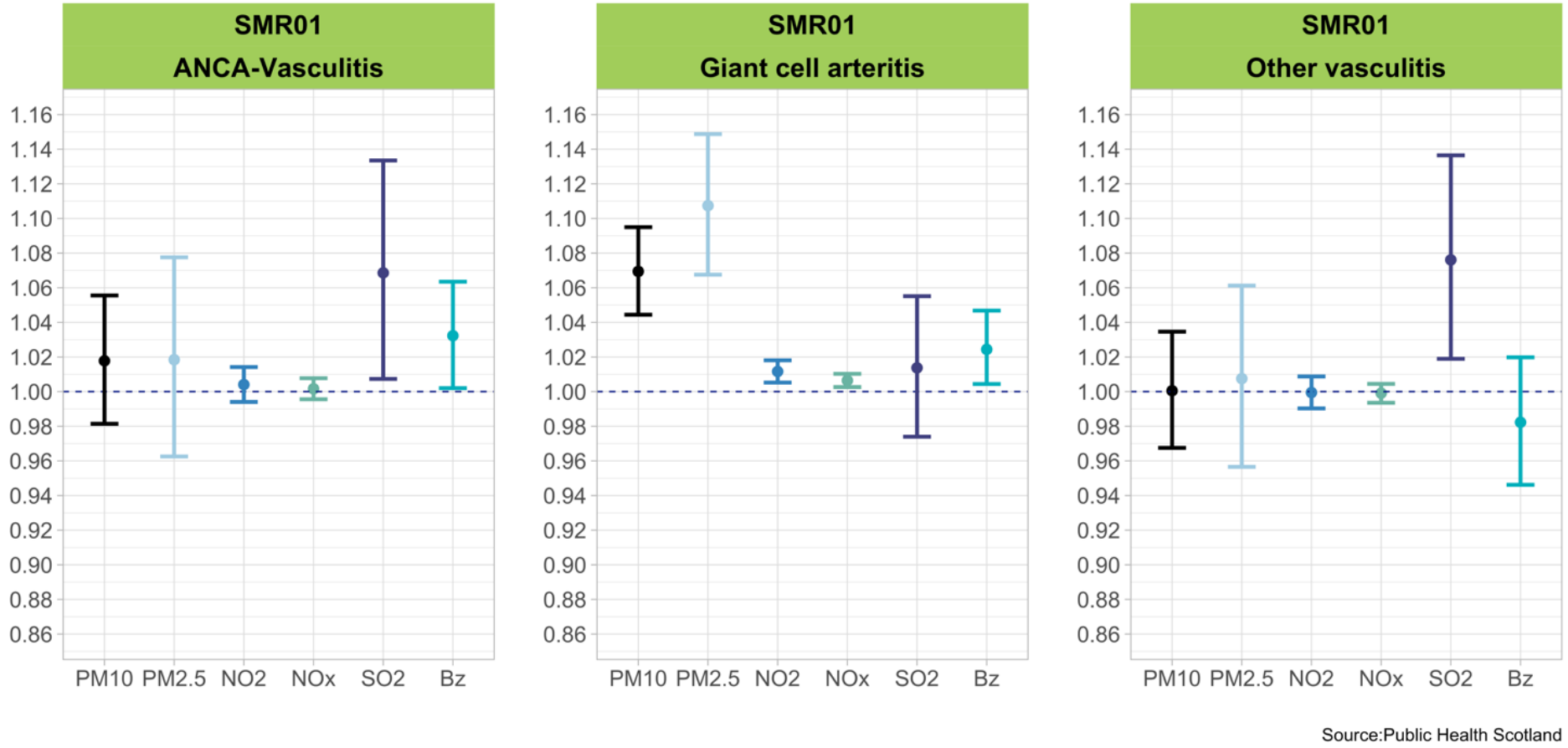



**Figure 6.3**: Odds ratio plot depicting the association between long-term air pollution exposure (2001-2020) and the odds of ANCA-Associated vasculitis, (A), Giant cell arteritis (B) and other vasculitis (C) in the VOICES Study

**Table 6.2:** Association between long term air pollution exposures and vasculitis in the VOICES Study

| | | Unadjusted | | | | | Adjusted [a] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Vasculitis phenotypes** | **Pollutant** | **Odds ratio** | **95% Confidence interval** | | **p-value** | **p-FDR** | **Odds ratio** | **95% Confidence interval** | | **p-value** | **p-FDR** |
| ANCA-Vasculitis | PM10 | 0.993 | 0.962 | 1.024 | 0.654 | 0.654 | 1.018 | 0.981 | 1.056 | 0.342 | 0.63 |
| | PM2.5 | 0.981 | 0.936 | 1.029 | 0.434 | 0.558 | 1.018 | 0.963 | 1.078 | 0.526 | 0.728 |
| | NO2 | 0.996 | 0.988 | 1.004 | 0.352 | 0.558 | 1.004 | 0.994 | 1.014 | 0.427 | 0.699 |
| | NOx | 0.997 | 0.992 | 1.002 | 0.253 | 0.506 | 1.002 | 0.996 | 1.008 | 0.594 | 0.764 |
| | SO2 | 1.048 | 0.991 | 1.109 | 0.101 | 0.260 | 1.069 | 1.007 | 1.134 | 0.028 | 0.072 |
| | Bz | 1.014 | 0.982 | 1.045 | 0.418 | 0.558 | 1.032 | 1.002 | 1.064 | 0.036 | 0.081 |
| Giant cell arteritis | PM10 | 1.073 | 1.051 | 1.095 | <0.001 | <0.001 | 1.069 | 1.044 | 1.095 | <0.001 | <0.001 |
| | PM2.5 | 1.106 | 1.071 | 1.141 | <0.001 | <0.001 | 1.107 | 1.068 | 1.149 | <0.001 | <0.001 |
| | NO2 | 1.013 | 1.008 | 1.0184 | <0.001 | <0.001 | 1.012 | 1.005 | 1.018 | <0.001 | <0.001 |
| | NOx | 1.007 | 1.004 | 1.011 | <0.001 | <0.001 | 1.006 | 1.003 | 1.010 | 0.001 | 0.005 |
| | SO2 | 1.022 | .0.985 | 1.061 | 0.248 | 0.506 | 1.014 | 0.974 | 1.055 | 0.504 | 0.728 |
| | Bz | 1.034 | 1.015 | 1.053 | <0.001 | <0.001 | 1.024 | 1.004 | 1.047 | 0.017 | 0.051 |
| Other vasculitis | PM10 | 1.012 | 0.983 | 1.042 | 0.977 | 0.558 | 1.001 | 0.968 | 1.035 | 0.977 | 0.977 |
| | PM2.5 | 1.019 | 0.974 | 1.065 | 0.414 | 0.558 | 1.008 | 0.957 | 1.061 | 0.777 | 0.874 |
| | NO2 | 1.003 | 0.995 | 1.01 | 0.522 | 0.626 | 0.999 | 0.990 | 1.009 | 0.912 | 0.966 |
| | NOx | 1.001 | 0.996 | 1.006 | 0.621 | 0.654 | 0.999 | 0.994 | 1.004 | 0.758 | 0.874 |
| | SO2 | 1.096 | 1.041 | 1.153 | 0.001 | 0.003 | 1.076 | 1.019 | 1.136 | 0.008 | 0.029 |
| | Bz | 0.991 | 0.959 | 1.023 | 0.574 | 0.646 | 0.982 | 0.946 | 1.020 | 0.35 | 0.63 |

[a.] adjusted for age, sex, ethnicity, SIMD, and population density

### 6.3.3. Assessing geographic differences in air pollution and the association with vasculitis in Scotland

The effects of air pollution on vasculitis cases were pronounced in those from rural areas. Figure 6.4 provides a summary of the geographic variation of the effect of long-term air pollution exposures on vasculitis. Overall, the stratified analyses shows that the effects of all air pollutants assessed here were significantly associated with AAV in those from rural areas. For the rural population, these four pollutants $PM_{10}$, $PM_{2.5}$, $SO_2$ and Benzene stood out to be significantly associated with AAV and GCA. For example, the effects of $PM_{2.5}$ and $SO_2$ in rural areas were linked with 18% (adjusted odds ratio: 1.18, 95% CI: 1.050 – 1.32) and 16%( adjusted odds ratio: 1.16, 95% CI: 1.03 – 1.30) increase odds of AAV (Figure 6.3A). Similarly, a 1 µg per cubic meter higher average levels of benzene in rural areas was associated with 16% ( adjusted odds ratio: 1.16, 95% CI: 1.05 – 1.28) increased odds of AAV. Similar effects were seen for GCA and $PM_{2.5}$, $SO_2$ and benzene (Figure 6.3B). Effects of nitrogen dioxide and nitrogen oxide were moderate but also statistically significant in populations from rural areas. These effects ranged from 3% to 6% for both AAV and GCA, including other vasculitis. A 1 µg per cubic meter higher average level of $NO_2$ over the period of 2001 to 2020 was associated with 5.5% ( adjusted odds ratio: 1.055, 95% CI: 1.029 – 1.082) and 5.6% ( adjusted odds ratio: 1.056, 95% CI: 1.038 – 1.075) increased odds of AAV and GCA among those residing from rural areas respectively. Similarly, 1 µg per cubic meter higher average level of NOx over the same period was associated with 5.5% ( adjusted odds ratio: 1.055, 95% CI: 1.029 – 1.082) and 5.6% ( adjusted odds ratio: 1.056, 95% CI: 1.038 – 1.075) and increased odds of AAV and GCA in those rural population. Lastly, the effects of $PM_{10}$, $PM_{2.5}$, $SO_2$ and Benzene remained statistically significant after correcting for multiple comparisons.

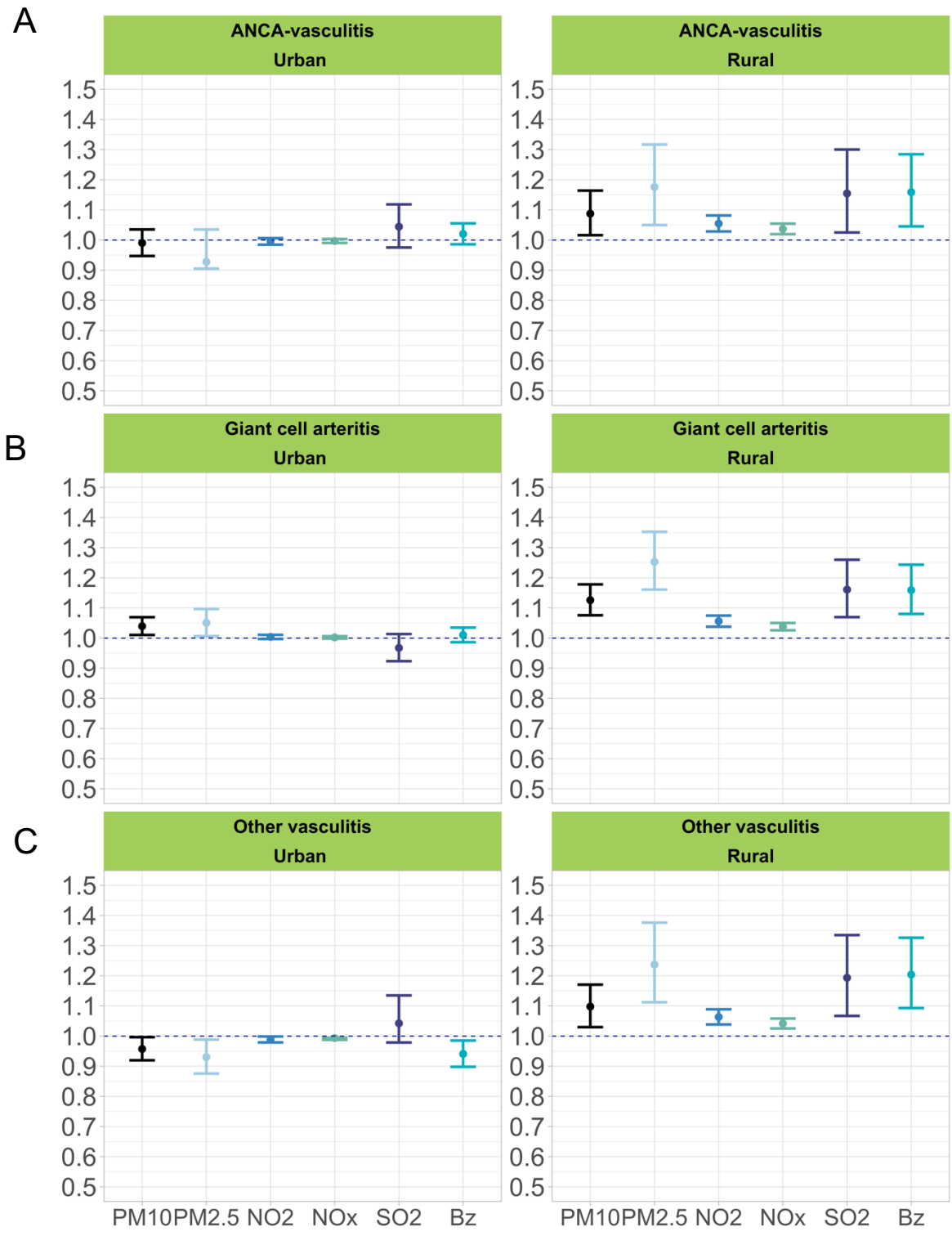



**Figure 6.4:** Stratified analyses showing the relationship between air pollution and ANCA-Associated vasculitis across different health boards in Scotland

**Table 6.3:** Summary of the stratified analyses showing the effects air pollution between urban and rural populations in Scotland

| | | | Unadjusted | | | | | Adjusted [b] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Outcome** | **Stratified** | **Pollutant** | **Odds ratio** | **95% Confidence interval** | | **p-value** | **p(FDR)** | **Odds ratio** | **95% Confidence interval** | | **p-value** | **p(FDR)** |
| ANCA-vasculitis | Urban | $PM_{10}$ | 0.965 | 0.927 | 1.006 | 0.092 | 0.114 | 0.990 | 0.947 | 1.035 | 0.664 | 0.664 |
| | Rural | $PM_{10}$ | 1.11 | 1.043 | 1.188 | 0.001 | 0.002 | 1.088 | 1.016 | 1.164 | 0.015 | 0.027 |
| | Urban | $PM_{2.5}$ | 0.932 | 0.875 | 0.993 | 0.029 | 0.044 | 0.928 | 0.905 | 1.035 | 0.341 | 0.384 |
| | Rural | $PM_{2.5}$ | 1.228 | 1.101 | 1.369 | 0.000 | 0.000 | 1.176 | 1.050 | 1.317 | 0.005 | 0.011 |
| | Urban | $NO_2$ | 0.990 | 0.979 | 1.00 | 0.050 | 0.067 | 0.995 | 0.985 | 1.006 | 0.417 | 0.429 |
| | Rural | $NO_2$ | 1.061 | 1.036 | 1.088 | 0.000 | 0.000 | 1.055 | 1.029 | 1.082 | 0.000 | 0.000 |
| | Urban | $NO_x$ | 0.994 | 0.987 | 0.999 | 0.042 | 0.058 | 0.997 | 0.991 | 1.004 | 0.365 | 0.398 |
| | Rural | $NO_x$ | 1.041 | 1.024 | 1.058 | 0.000 | 0.000 | 1.037 | 1.019 | 1.054 | 0.000 | 0.000 |
| | Urban | $SO_2$ | 1.033 | 0.965 | 1.105 | 0.355 | 0.376 | 1.044 | 0.975 | 1.118 | 0.214 | 0.275 |
| | Rural | $SO_2$ | 1.188 | 1.060 | 1.332 | 0.003 | 0.005 | 1.155 | 1.025 | 1.300 | 0.018 | 0.031 |
| | Urban | Bz | 1.009 | 0.972 | 1.046 | 0.654 | 0.654 | 1.020 | 0.986 | 1.055 | 0.250 | 0.310 |
| | Rural | Bz | 1.175 | 1.063 | 1.299 | 0.002 | 0.004 | 1.159 | 1.045 | 1.285 | 0.005 | 0.011 |
| Giant cell arteritis | Urban | $PM_{10}$ | 1.064 | 1.036 | 1.092 | 0.000 | 0.000 | 1.039 | 1.010 | 1.069 | 0.008 | 0.016 |
| | Rural | $PM_{10}$ | 1.124 | 1.074 | 1.176 | 0.000 | 0.000 | 1.126 | 1.076 | 1.178 | 0.000 | 0.000 |
| | Urban | $PM_{2.5}$ | 1.085 | 1.041 | 1.130 | 0.000 | 0.000 | 1.051 | 1.006 | 1.097 | 0.024 | 0.038 |
| | Rural | $PM_{2.5}$ | 1.241 | 1.151 | 1.338 | 0.000 | 0.000 | 1.253 | 1.161 | 1.352 | 0.000 | 0.000 |
| | Urban | $NO_2$ | 1.008 | 1.002 | 1.015 | 0.013 | 0.022 | 1.004 | 0.997 | 1.011 | 0.280 | 0.336 |
| | Rural | $NO_2$ | 1.054 | 1.036 | 1.072 | 0.000 | 0.000 | 1.056 | 1.038 | 1.075 | 0.000 | 0.000 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Urban | $NO_x$ | 1.005 | 1.001 | 1.009 | 0.019 | 0.030 | 1.002 | 0.998 | 1.006 | 0.294 | 0.341 |
| | Rural | $NO_x$ | 1.036 | 1.024 | 1.048 | 0.000 | 0.000 | 1.038 | 1.026 | 1.050 | 0.000 | 0.000 |
| | Urban | $SO_2$ | 0.974 | 0.932 | 1.019 | 0.261 | 0.294 | 0.967 | 0.923 | 1.013 | 0.161 | 0.215 |
| | Rural | $SO_2$ | 1.154 | 1.064 | 1.251 | 0.001 | 0.002 | 1.161 | 1.069 | 1.260 | 0.000 | 0.000 |
| | Urban | Bz | 1.021 | 0.999 | 1.043 | 0.064 | 0.082 | 1.010 | 0.986 | 1.035 | 0.405 | 0.429 |
| | Rural | Bz | 1.146 | 1.069 | 1.229 | 0.000 | 0.000 | 1.159 | 1.080 | 1.243 | 0.000 | 0.000 |
| Other vasculitis | Urban | $PM_{10}$ | 0.985 | 0.948 | 1.023 | 0.427 | 0.439 | 0.957 | 0.920 | 0.996 | 0.032 | 0.044 |
| | Rural | $PM_{10}$ | 1.101 | 1.034 | 1.173 | 0.003 | 0.005 | 1.098 | 1.030 | 1.171 | 0.004 | 0.010 |
| | Urban | $PM_{2.5}$ | 0.964 | 0.909 | 1.022 | 0.216 | 0.259 | 0.930 | 0.876 | 0.988 | 0.019 | 0.031 |
| | Rural | $PM_{2.5}$ | 1.245 | 1.121 | 1.384 | 0.000 | 0.000 | 1.237 | 1.112 | 1.376 | 0.000 | 0.000 |
| | Urban | $NO_2$ | 0.994 | 0.985 | 1.004 | 0.252 | 0.293 | 0.989 | 0.979 | 0.999 | 0.025 | 0.038 |
| | Rural | $NO_2$ | 1.066 | 1.041 | 1.092 | 0.000 | 0.000 | 1.063 | 1.038 | 1.089 | 0.000 | 0.000 |
| | Urban | $NO_x$ | 0.997 | 0.991 | 1.003 | 0.280 | 0.305 | 0.993 | 0.988 | 0.994 | 0.030 | 0.043 |
| | Rural | $NO_x$ | 1.043 | 1.027 | 1.060 | 0.000 | 0.000 | 1.042 | 1.025 | 1.058 | 0.000 | 0.000 |
| | Urban | $SO_2$ | 1.077 | 1.013 | 1.146 | 0.017 | 0.028 | 1.042 | 0.979 | 1.135 | 0.002 | 0.005 |
| | Rural | $SO_2$ | 1.203 | 1.078 | 1.343 | 0.001 | 0.002 | 1.193 | 1.067 | 1.335 | 0.002 | 0.005 |
| | Urban | Bz | 0.954 | 9.914 | 0.996 | 0.034 | 0.049 | 0.941 | 0.898 | 0.985 | 0.010 | 0.019 |
| | Rural | Bz | 1.216 | 1.104 | 1.339 | 0.000 | 0.000 | 1.204 | 1.093 | 1.326 | 0.000 | 0.000 |

[b] adjusted for age, sex, ethnicity, SIMD and population density

## 6.4.Discussion

In this large nationwide study using routine health data from the Scottish morbidity record, we observed a significant association between $SO_2$ and $PM_{2.5}$ exposure over the period between 2001 to 2020 and onset of AAV and GCA in Scotland. Specifically, $SO_2$ showed a significant association with AAV before adjustment for multiple comparison, while $PM_{2.5}$ was significantly associated with GCA before and after adjusting for multiple comparison. This variation in effects of air pollution is an important addition to findings seen from UK Biobank. From UKB, we observed that the effects of $SO_2$ were 6.4% (odds ratio: 1.064) higher for the onset of vasculitis. In SMR01, we see that effects of $SO_2$ are 6.9% (odds ratio: 1.069) for the onset of AAV and 7.6% (odds ratio: 1.076) for other vasculitis even with limited adjustment for confounders that were accessible from the routine health records. For GCA, we observed a novel association with $PM_{10}$ and $PM_{2.5}$, suggesting that the association between air pollution and vasculitis may vary between small vessel and large vessel vasculitis.

The consistent results seen in SMR01 and UKB suggests that the effects of $SO_2$ may be attributable to small vessel vasculitis, specifically AAV. Sulphur dioxide also showed to be associated with other autoimmune disease like rheumatoid arthritis and systemic lupus erythematosus. This association between $SO_2$ and autoantibody mediate diseases indicates that this pollutant is potential trigger of B cell immunity. Studies from mouse models have shown that $SO_2$ can induce an autoinflammatory immune response through the activation of Toll-like receptor-4 (TLR-4) or nuclear factor -kB (NF-kB) pathways. This activation leads to increased production of reactive oxygen species (ROS) and an amplification of $Th_2$ cytokines (333). Evidence from human studies show that $SO_2$ is associated with 170 differentially expressed genes enriched in pathways involved in leukocyte migration during chronic inflammation (334) and B cell migration through CXCR chemokine activity and G-protein couple receptors (335). Possible mechanism associated with GCA and the particulate matters ($PM_{10}$ and $PM_{2.5}$ ) are not clear. A recent study by Ural et al.,2022 could provide insight into how particular matter can alter

immune response in elderly population and possibly predisposing them to GCA. For example, the study Ural et al.,2022 showed that aging was directly linked with cumulative exposure and concentration of $PM_{2.5}$ in the lymph nodes. The authors observed that $PM_{2.5's}$ were specifically contained within $C68^{+}CD169^{-}$ macrophages which exhibited decreased function and altered cytokine production (336). These results reported from this chapter and previous chapter provide the first ever evidence on the effects of air pollution on the different forms of vasculitis and suggests that pathways by which air pollution impacts onset of vasculitis may differ depending on the vessel size.

The other major finding from this study is that there are varying effects of air pollution on the risk of vasculitis based on urban and rural geographies. These effects were consistent with the results from UK Biobank. In UKB, stratified analyses based on population density showed that rural populations had 18% higher risk of developing vasculitis (odds ratio:1.18) due to $SO_2$ compared with urban population (odds ratio:1.05). For the SMR01, the effect of $SO_2$ on AAV and other vasculitis were also found to be higher for rural population. Those from rural areas had 16% % higher odds of developing AAV (odds ratio: 1.16) compared with urban population (AAV odds ratio: 1.044). Together, these results showed that individuals from rural areas may be vulnerable to air pollution in relation to vasculitis

While the results reported here are novel and important for the field of vasculitis, they should be interpreted in light of the following limitations. First, the use of ICD codes to classify AAV phenotypes may not be sensitive. Cook et al., 2022 recently validated the diagnoses of AAV in a clinical registry derived using administrative health data in the US. The authors observed that patients coded as having GPA which is mostly associated with proteinase-3 (PR-3-ANCA) actually were MPA (>50% cases were MPO-ANCA+) (337). To minimise this, MPA, GPA and EGPA cases were combined to define AAV and to provide power needed for environmental study. Other limitations were related to accessibility of study variables, accuracy of exposure, and potential confounder. Routine data do not collect individual level risk factors, such as smoking. Smoking is an important confounder of AAV and the exposure of air pollution may also not

reflect true exposure given that most people spend majority of time indoors in Scotland. Regardless of said limitations, UKB results suggested that adjusting for smoking may not have significant impact in the direction of effect of air pollution on vasculitis.

Lastly, it is worth noting while there are several limitations to consider, this study have several advantages. First, the study data provided ample power and spatial coverage to detect important association between between the different air pollution exposures and the onset of vasculitis. This statistical power is particularly useful in environmental studies where exposure is highly prevalent but variable across whole population, often leading to small effect estimates. The study data also covered complete population of Scotland and included all possible cases of vasculitis diagnosed between 1996 to 2020 based on their clinical phenotypes. The consistent use of the same method across the UKB and SMR01 study also allowed for appropriate comparison of results, providing strong evidence on the effect of air pollution and the incidence of vasculitis.

# Chapter 7 – Understanding the temporal and seasonal variation in ANCA-associated vasculitis in Scotland

## 7.1. Overview

The goal of this chapter is to address the third objective of this thesis, which is to understand the temporal and seasonal cycles of vasculitis in the UK. This is done using routine health data from the Scottish morbidity record under the VOICES project. The UKIVAS could not be used for the seasonal analyses as recruitment to the registry is dependent on nurse availability and is therefore subject to substantial temporal bias (338) and would not serve as a good validation dataset. As such the SMR01 was specifically used to assess the seasonal variation in ANCA-Associated vasculitis in Scotland. This dataset was used due to its better surveillance and national coverage of vasculitis patients in Scotland.

The seasonal analyses reported here will add to the growing body of literature reporting on the temporal cycles and variation of vasculitis. As demonstrated in the review from chapter 2, currently, there is mixed evidence about the seasonal nature of AAV in studies from both the northern and southern hemisphere (130,157,159,160,217,229,339). Half of these studies (57%, n=4 out of 7 studies) have found no seasonal variation in AAV. The rest of the studies vary based on the definition used to describe the onset of disease. For example, two of the three studies reporting a seasonal variation in AAV used the date of diagnosis as a measure of the disease onset. In these studies, the incidence of AAV was found to increase during warmer seasons of the year (Spring and summer, p-value <0.01) (229,339). The only study to report higher incidence of AAV in winter used the date of first symptom onset to define disease onset (130). This study was also the only one to use routine health data, mainly hospital record with a wide national population coverage. The rest of the studies used patient data from a single specialist centre or a renal registry. Majority

these studies were limited in size and had a sample sizes ranging from 29 cases to 339 cases.

The study reported in this chapter will address these limitation and with its large n and temporal coverage from a nationally representative dataset covering over 1600 AAV patients diagnosed between 1996 and 2020 in Scotland

### 7.1.1. Study population

As mentioned above, the seasonality of ANCA-associated vasculitis was assessed using admission data from the Scottish Morbidity Records (SMR01) as it has the most contemporary national coverage of people living with vasculitis across Scotland and have the longest temporal coverage with earliest diagnoses that dates as early as the 1990s. The UKIVAS could not be used as recruitment into the registry is dependent on nurse availability and is subject to temporal bias (338) and would thereby not serve as a good validation dataset for the seasonal analyses and has limited validity in the case representing the seasonality of vasculitis

### 7.1.2. Scottish Morbidity Record (SMR01)

As mentioned in chapter 3 and 6, the SMR01 is a national administrative health dataset with records on all inpatient and day case hospital discharges from non-obstetric and non-psychiatric hospitals across Scotland. It includes episodic data collected at hospital admissions and discharge, and with the purpose of providing clinical service, including monitoring wait times and duration of care. For instance, a single admission may contain multiple episodes if a patient is transferred to another consultant and speciality. The main condition of treatment and up to five other conditions may be recorded for each episode and is coded using the International Classification of Disease codes (ICD-10). It is possible for different episodes within one admission to have different main conditions. For this chapter, AAV patients were identified in either the main or one of the five other conditions categories in the SMR01 using relevant diagnostic codes of vasculitis (M31.3 for granulomatosis with polyangiitis (GPA), M31.7 for microscopic polyangiitis (MPA) and M30.1 eosinophilic granulomatosis with

polyangiitis (Churg-Strauss/EGPA)). Figure 7.1 shows a graphical summary of the age distribution of these patients grouped as AAV.

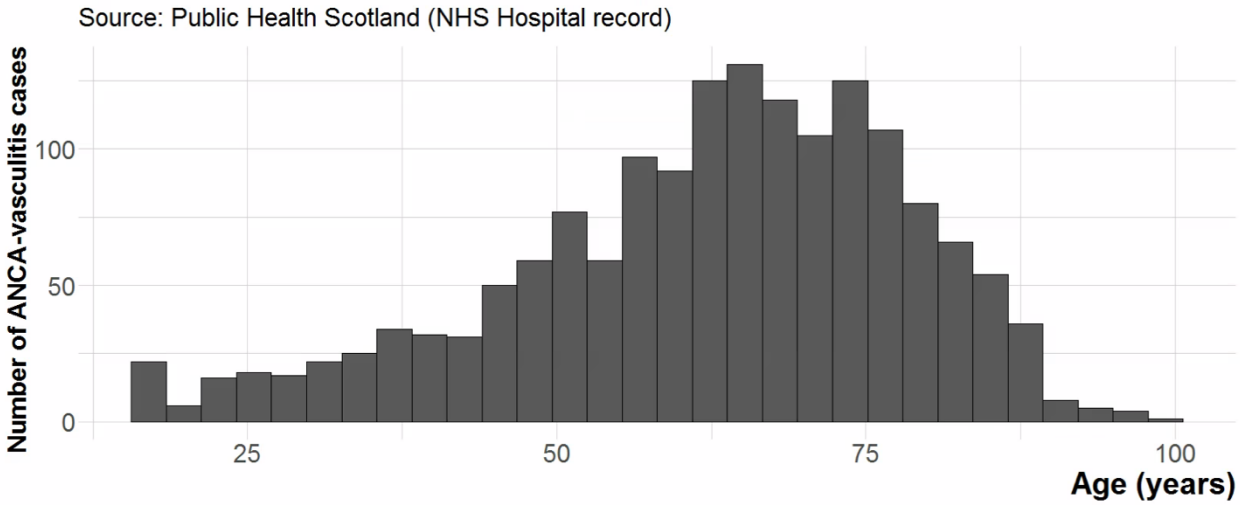



**Figure 7.1:** The age distribution of patients with ANCA-associated vasculitis between 1996 to 2020 in Scotland (SMR01)

### 7.1.3. Selection criteria

Participants were included in the final analyses if they had a complete date of diagnosis (DMY).All identified patients with AAV had a full date of diagnosis and the number of diagnoses were equally distributed when aggregated based days of the month (1$^{st}$ to 31$^{st}$) (**Figure 7.2**). This distribution in monthly diagnoses was used to guide the selection criteria of patients included in the analyses and it served as quality check with regards to how the dates were recorded in the SMR01 and to detect any data capturing errors. Overall, each day of the month had less than 4% of AAV cases for all the diagnoses captured between 1996 to 2020.

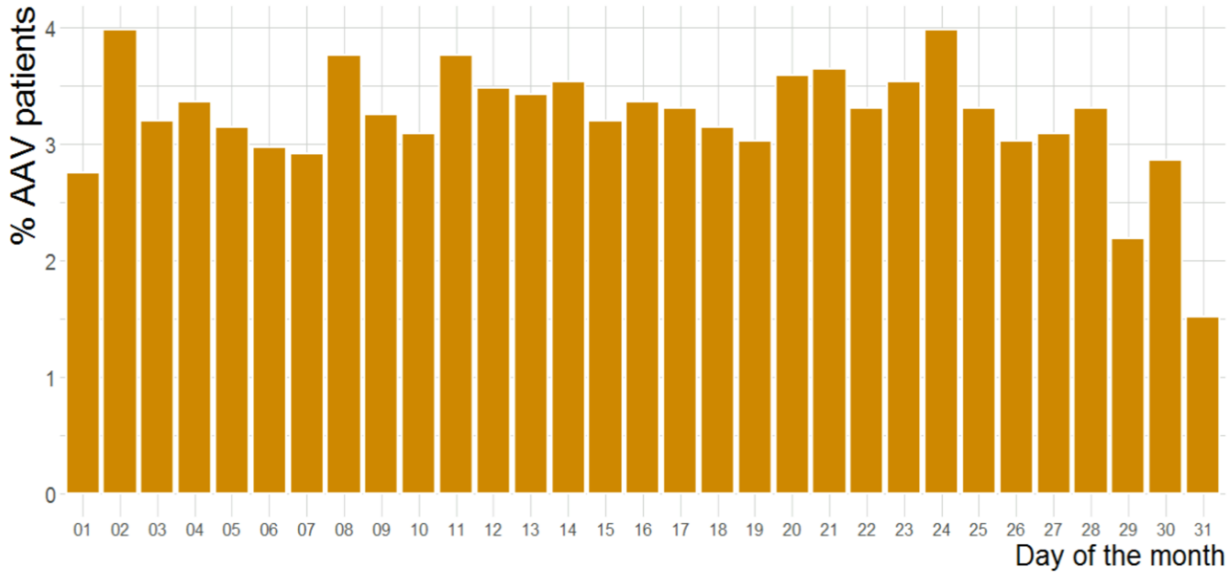



**Figure 7.2:** Distribution of AAV diagnoses from 1996 to 2020 in the SMR01. The dates of diagnosis were aggregated according to their corresponding day of the month (1st to 31st). This exploratory analysis served as quality check, to assess for outliers or any error

### 7.1.4. Statistical analysis

#### 7.1.4.1. Periodicity and interannual variation in AAV onset

The temporal and seasonal cycles of AAV occurrences between 1996 to 2020 were assessed using the decomposition time series method, first by using the Seasonal and Trend decomposition using Loess approach (STL) and non-stationary cosinor (NSC) method for the sensitivity analysis. The aim of the time series decomposition is to separate noise from the signa. Noise may arise due to random variance or time associated changes due to changes in disease classification or diagnosis. This also considers other random and focuses on the signal (if any). After removing noise from the time series, the signal should reflect the 'real' trend (non-stationary mean) in the data including its variability.

Specifically, for the STL decomposition, a loess smooth function is used to model the trend and seasonal effects using numerical method that are based on the data itself and not on any mathematical assumption or specific probability about the distribution of the time series (340).

This is done in two parts: signal and noise

$$Y_t = X_t + e_t \qquad t = 1, .....n \qquad \text{(Equation 4.1)}$$

where $X_t$ is the signal part of the series, and the Greek letter $e_t$ represents the noise (or residuals). The n is the total length of the time series or sample size. In practice, the STL algorithm is achieved using four steps, where the dependent variable *Y* and independent variable *X* are used to complete the loess smooth functions:

i. First, the values representing $X_t$ define the window period which is $X_t$ - *h* to $X_t$ + *h* for some constant *h*. The pair, the dependent and independent variables, are labelled as *Y'* and *X'*, respectively.

ii. Weights for every single point in the window $wj = f(x' - xt)$ are then calculated. These weights describe the distance between each point and the central point (*Xt)*, with values further from *Xt* being given less weight.

iii. A regression model is then fitted $Y' = f(X')$ using these weights *w*. The model is often linear $f(u) = ßo + ß_1u$ but also can be a quadratic spline $f(u) = ß_0 + ß_2u^2$. The $f(u) = ß_0$ of the formula represents *moving averages*, with $ß_1$ being function of a linear model and $ß_2u$ being a function of a spline model.

iv. Lastly, the fitted values are estimated using $st = y'_t$ at $x_t$ function for a given window

These four steps are repeated for a range of different values for *X* to build a smooth seasonal curve. For monthly data covering long period of time with a seasonal pattern, the STL algorithm groups each month and applies a loess

smooth function to each group or year. The results for every January are smoothed, then again for every February, and so on. The widow ($h$) is the number of neighbouring years included in the smooth. Selecting $h$=1 means only one neighbouring year is used, whereas $h$=2 uses two years from either side. The recommended size of the smooth window is $w$= $2h$+1 which estimates the long-term pattern in seasonality based on each month. The smoothing is done in two loops, first by iterating the seasonal ($S_t$), then the trend smoothing ($\mu_t$). This is followed by calculating the effect of noise or residuals in the series ($e_t$). For the noise in the data, the residuals are assumed to be uncorrelated, with the mean being zero and variance being constant. The 'real' trend in the data is calculated by subtracting the seasonal and trend component from the time series to give a summary of the residuals (noise/ remainder) (Figure 7.3) – see below figure for the concept behind the STL algorithm.

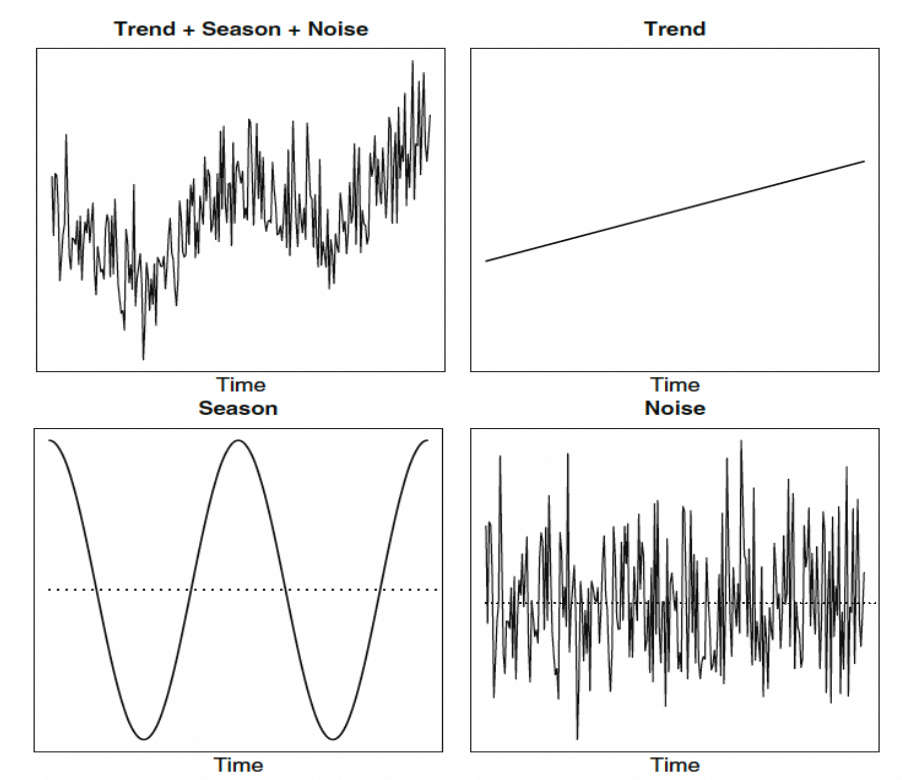



**Figure 7.3:** Decomposing a time series to capture the trend, season, and noise. The Y scales for seasonality and noise are different to reflect their corresponding estimated values for each component.

The remainder values represent the amount of noise in the data. The noise in the data represent periods with excess cases (outliers) which can be relatively large or negative compared to other remainder values. Values close to zero give indication that the seasonal and trend components are accurate in modelling the time series, while large values indicate the presence of noise. Advantages of using this approach is that it's able to separate short-term changes in the trend from seasonality and vice versa.

For the sensitivity analysis, a non-stationary cosinor model was used. This approach uses cubic spline for the trend, and a non-stationary seasonal pattern which are defined as:

$$S_t = A_t \cos(w_t - P_t), \quad t=1,\ldots\ldots, n, \qquad \text{(Equation 4.2)}$$

Where $2wt = 2\pi ft$ and *ft* is the fraction of time for individual dates (e.g.: date of diagnosis). This model is parametric as it is constraint to sinusoidal pattern. However, it is non-stationary because the amplitude (At) and phase (Pt) are dependent on time. Additionally, the flexibility in the non-stationary seasonal estimate is determined by the variance and the smoothing parameter and allows for multiple seasonal components to be added to fit two seasonal terms.

### 7.1.4.2.Seasonal variation in AAV onset

The seasonal analyses were informed by the literature reported in Chapter 2 which showed that most studies reporting on the seasonal variation in AAV use mainly non-parametric descriptive statistics and Poisson regression to compare to incidence of vasculitis across the different seasons of the year. Given the incidence of AAV first aggregated based on the month and seasons of diagnosis. This was examined and assessed for uniformity using one-way ANOVA and compared using a pairwise t-test with corrections for multiple testing. Seasons were defined using the meteorological definition and were divided into spring (March - May), Summer (June - August), Autumn (September - November), Winter (December - February)

**Table 7.1:** Meteorological Seasons (Northern hemisphere)

| | |
|---|---|
| Winter | December, January, February (91 days) |
| Spring | March, April, May (92 days) |
| Summer | June, July, August (91 days) |
| Autumn | September, October, November (91 days) |

Furthermore, a generalised additive model (GAM) was used to robustly model the seasonal variation in incidence of AAV based on days the year (365 days) and further provide moving confidence interval for each month and season of the year. The model can be summarised as follows:

$$Y_t = a + s(t, \lambda), \qquad t = 1,\ldots\ldots..,n$$

Where $s(t, \lambda)$ represents a smooth function of time $(t)$, controlled by the degrees of freedom $\lambda$, usually greater than 1, in this case 10. This is important because large $\lambda$ lead to more flexible (sinusoidal or bendy) functions, while smaller values lead to more linear functions that may underfit the data. The key assumption here is that the daily and monthly count events have an over-dispersed Poisson distribution. Furthermore, Index year with a cubic smoothing function ($\lambda$ = 10) was used to control for long-term variability in disease incidence factored by random function associated with time

## 7.2. Results

### 7.2.1. Study characteristics

The characteristics of AAV patients in Scotland are summarised in Table 7.2. The mean age was 60 years old (±15.96) with over half of the patients being male (52.4%). Most of the AAV patients were from urban areas (76%), with a fifth of the population (20.8%) being from the least deprived areas. 33% of all cases were diagnosed between 1996 and 2000 following the discovery ANCA testing in 1995 and its increased use in clinical settings to diagnose and classify patients with AAV (341,342). This is reflected with a sharp peak in monthly incidence of AAV between 1996 to 2002 (Figure 7.4). In fact, 33.4 % AAV population were diagnosed during this period, perhaps reflecting the lag effects in detection of previously undiagnosed cases (Table 7.2). The peak between 1996 and 2000 was followed by a stable 18-year sinusoidal pattern in AAV incidence that ranged from 21.3% to 23.8% of temporal cases between 2003 to 2020.

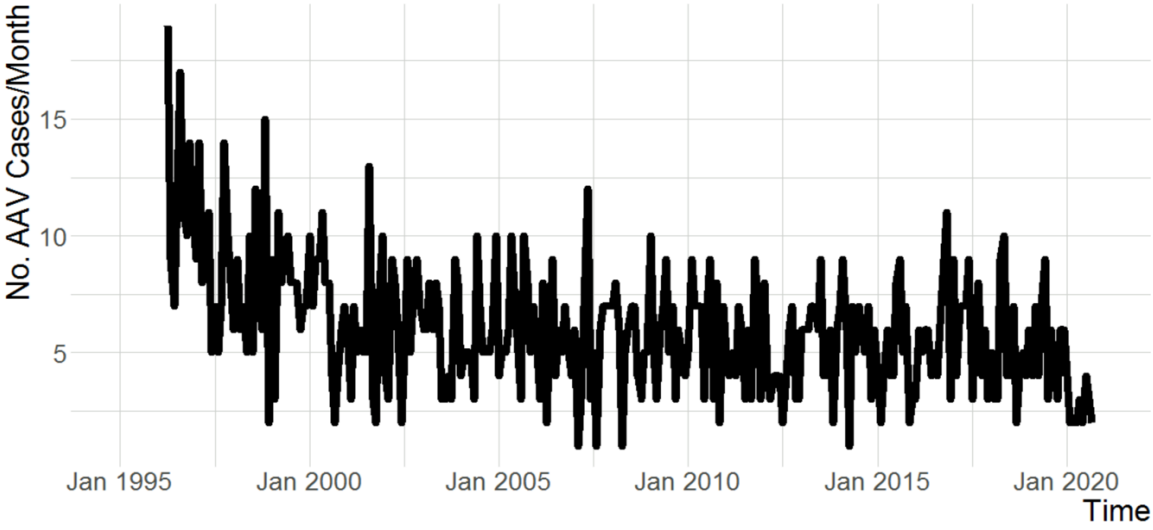



**Figure 7.4:** Number of ANCA-vasculitis cases in the SMR01 covering the period of 1996 to 2022.

**Table 7.2:** Characteristics of patients with ANCA-Vasculitis from the SMR01

| | SMR01 |
|---|---|
| | **AAV (n=1,622)** |
| Overall incidence | |
| Age, mean (SD) | 60 (±15.96) |
| Sex, n (%) | |
| Male | 850 (52.40%) |
| Follow up, median | 6.6 (2.40 – 14.34) |
| Rural/Urban (Scotland), n (%) | |
| Urban area | 1231 (76%) |
| Rural area | 387 (24%) |
| Scottish IMD, n (%) | |
| 1 (most deprived) | 290 (18%) |
| 5 (least deprived) | 350 (20.8%) |
| Period, n (% and Incidence per 100,000) | |
| Overall (1996 to 2020) | 1,622<br>29.7 cases per 100,000 |
| 1996 - 2002 | 542<br>(33.4%)<br>10.7 cases per 100,00 population |
| 2003 - 2008 | 345<br>(21.3%)<br>6.6 cases per 100,000 population |
| 2009 - 2014 | 386<br>(23.8%)<br>7.35 cases per 100,000 population |
| 2015 - 2020 | 349<br>(21.5%)<br>6.43 cases per 100,000 population |

### 7.2.2. Periodicity and interannual variation in vasculitis incidence

The overall trend in daily incidence of AAV shows an oscillating pattern with peaks and trough being seen, initially, every 2 years (1998-2000) then every 4 to 5 years between 2002 to 2020 (Figure 7.5A). Specifically, there were two major periodic cycles, consisting of two peaks and troughs first between 1999-2000 (1st peak) then 2017-2018 (2nd peak) and 2004 (1st trough) and 2013 (2nd trough). This variation is captured by the trend (non-stationary mean) which shows extreme interannual variation that ranges from 5 to 8 cases per month (Figure 7.5B). Prior to the first peak, there was higher-than-normal incidence of AAV with a remainder value of 2.5, potentially indicating noise and in the time series (Figure 7.5C). Similar patterns are also seen with later years. Seasonal estimates showed a relatively constant seasonal annual cycles in AVV cases over time. This variation follows a sinusoidal and non-stationary pattern, with sharp peaks in the second quarter of the year (April-June) and troughs in the last months of the year (October-December). Overall, the amount of noise in the time series represented by the remainder values preceded most of the major peaks in AAV incidence. This can be seen with high remainder values in 1998, that are elevated at an incidence peak of 1999-200. Other periods with high remainder values included 2002,2007, 2008, 2014, with incremental increases over time. Sensitivity analysis using non-stationary cosinor time series showed a similar pattern with increased mean in AAV being seen during the early years of ANCA testing and use(1996-2000), which levelled with later years (Figure 7.6).

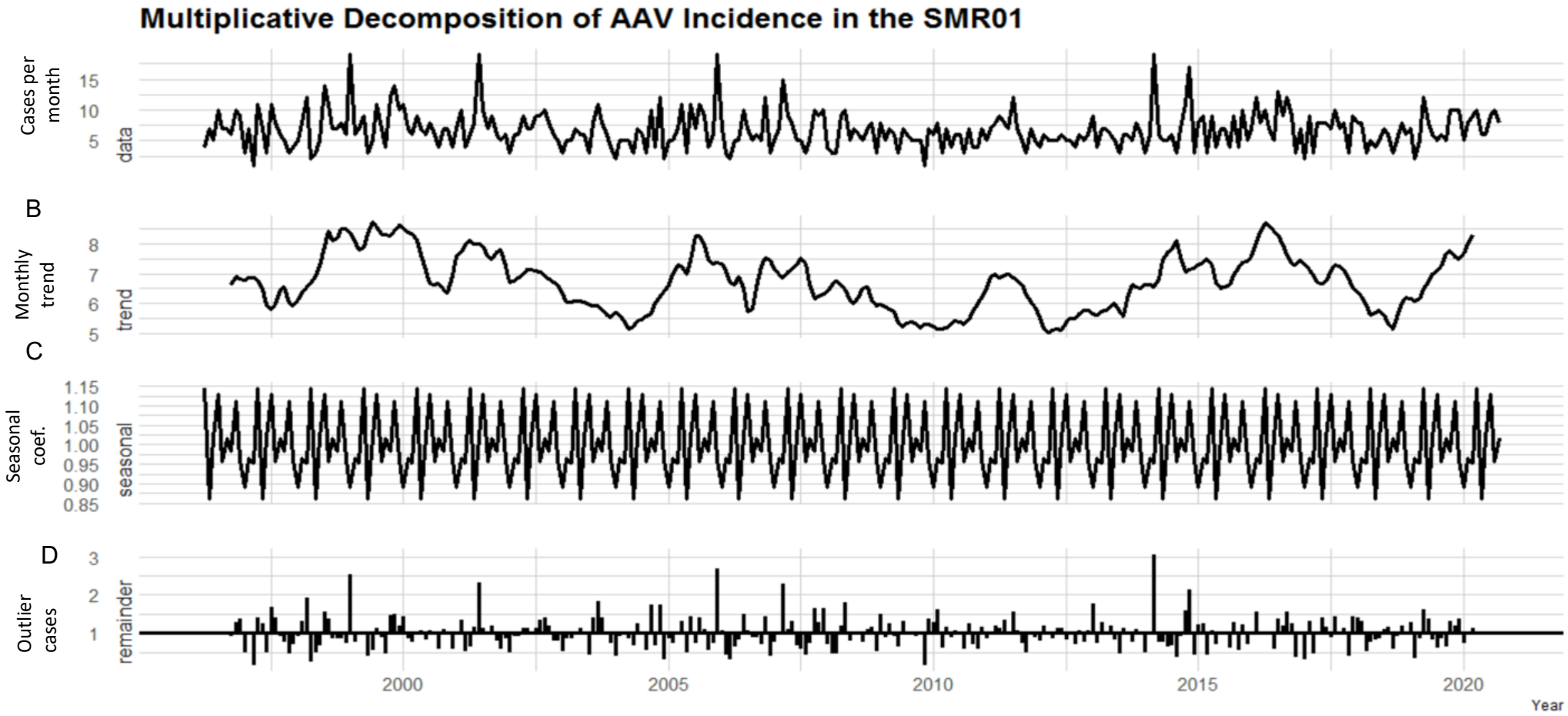



**Figure 7.5:** Results of the Seasonal and Trend decomposition time series using Loess (STL) of ANCA-vasculitis in SMR01.The figure summarises the absolute incidence of AAV (A) and includes the trend or periodic cycles (B), seasonal cycles (C) and overall noise (outliers)

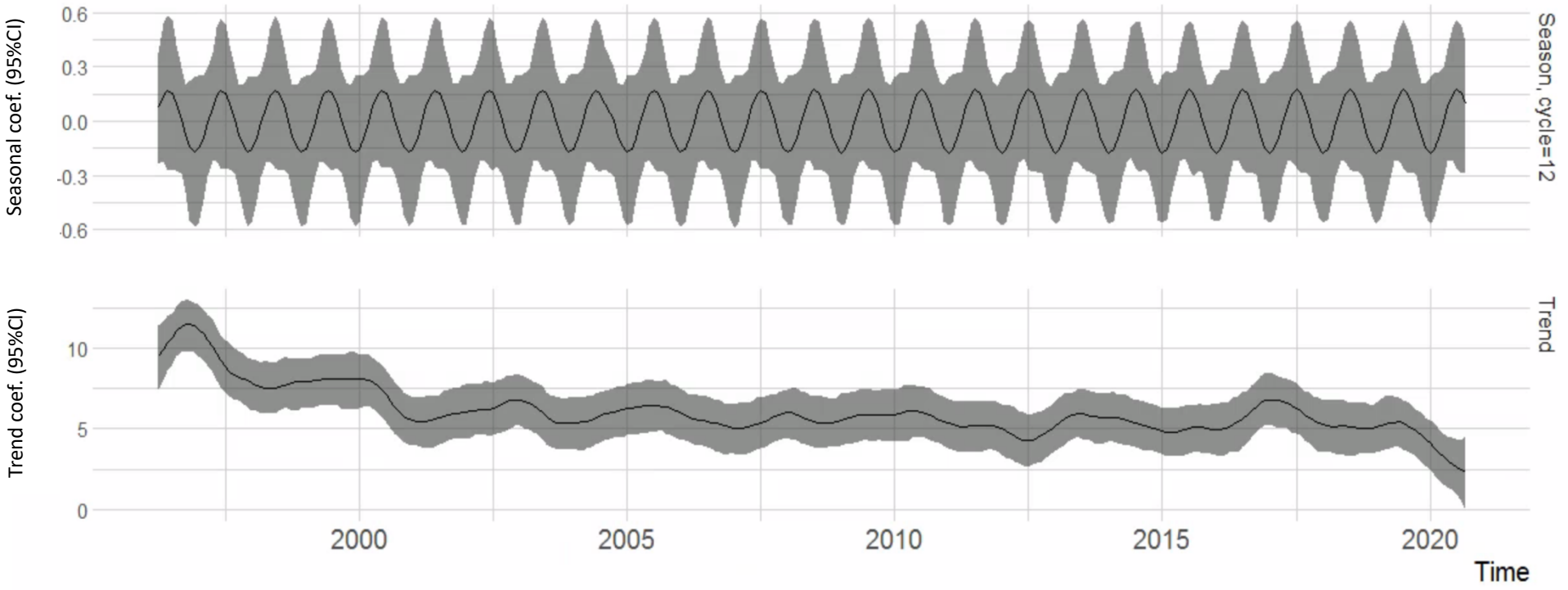



**Figure 7.6:** Sensitivity analysis using cubic spline function for trend and non-stationary seasonal analysis using smooth parameter $T_1$ =10, $T_2$ =52.The solid line represents the rolling mean in AAV cases, and the grey line are the 95% confidence interval.

### 7.2.3. Seasonal variation of AAV

Figure 7.7 provides a detailed summary of the daily, monthly, and seasonal variation in AAV incidence in Scottish Morbidity Records (SMR01). Overall, the daily incidence of AAV followed a sinusoidal pattern, with major and minor peaks being seen between the 10$^{th}$ and 15$^{th}$ of each month. Specifically, there were two peaks at day 190 and 330, with varying trend in peaks being seen in the middle (15$^{th}$) of each month of the year. These peaks were relatively consisted throughout the year, with little to no variation in between months of the year. When aggregated according to months and seasons of the year and compared using pairwise t-test corrected for multiple comparison using Bonferroni correction (pairwise p(FDR) = 0.66) showed that there was no statistically significant seasonal variation in AAV between 1996 to 2020 in the Scottish morbidity records

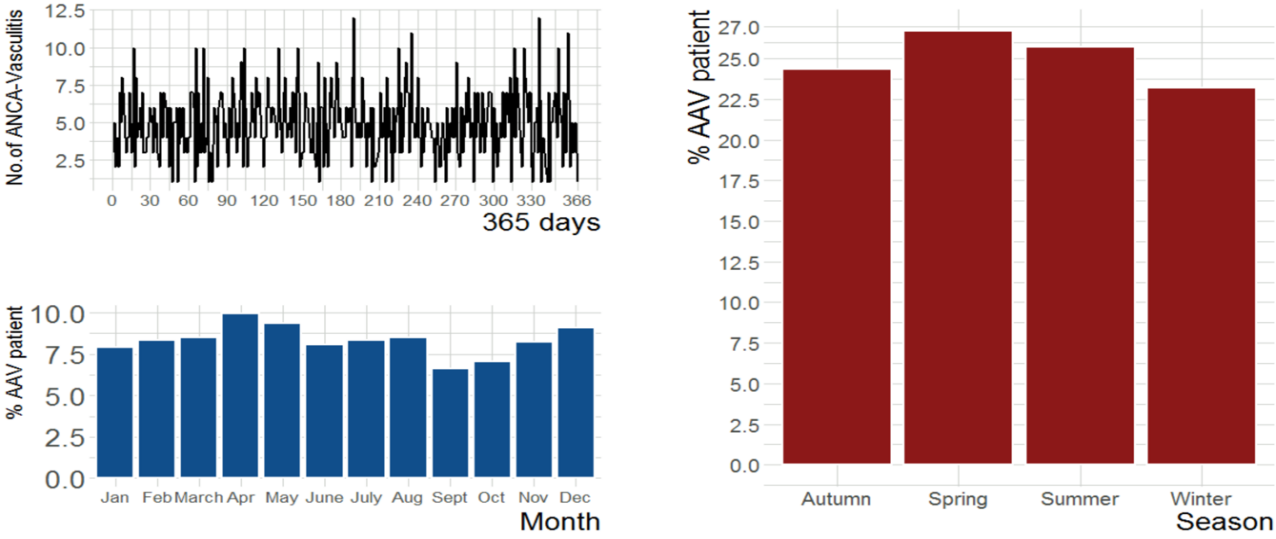



**Figure 7.7:** Seasonal variation of ANCA-Vasculitis in SMR01

Additionally, a generalised additive model (GAM) and Poisson regression were run to model seasonality of AAV and to provide additional sensitivity analysis that adds to the frequency comparison of seasonality reported in Figure 7.9. A smooth function ($S_t$) was applied to model the seasonal variation in AAV based on AAV annual count data aggregated based on days of the year (365 days)

(Figure 7.8). A 10 degrees of freedom was used to allow a close fit the seasonal data. The results of this modelling approach showed a small peak in incidence of AAV between days 90 to 120 of the calendar year. Result of the Poisson regression model showed a non-statistically significant monthly or seasonal variation in AAV as the confidence interval crosses the null (Figure 7.9)

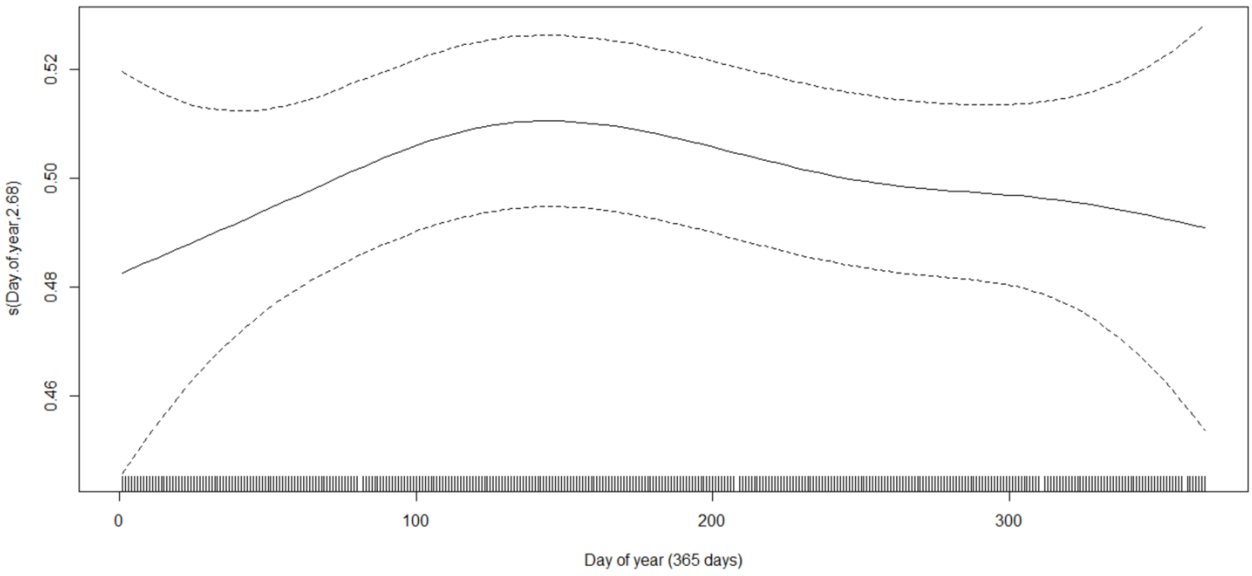



**Figure 7.8:** Generalised additive modelling results of the AAV seasonal distribution in the SMR01

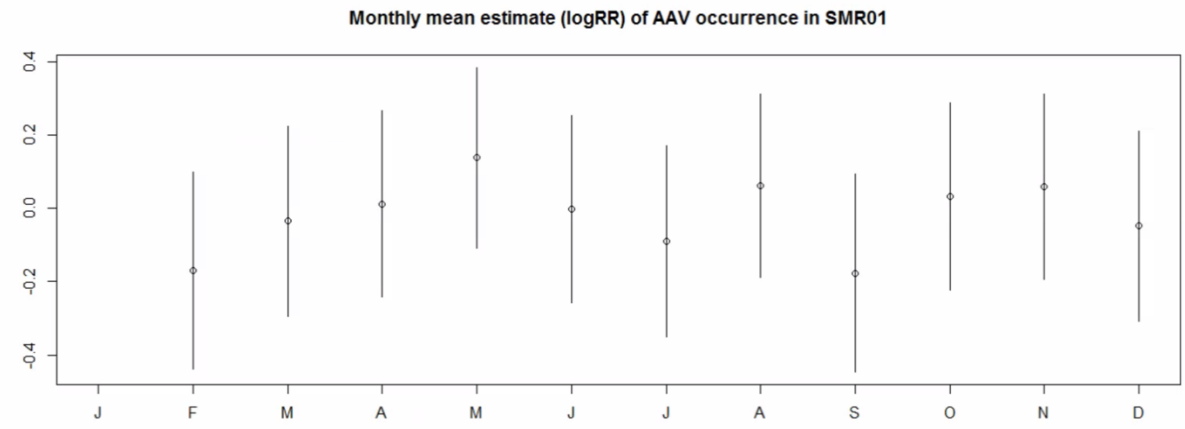



**Figure 7.9:** Monthly mean estimates of AAV incidence in SMR01 using an overdispersed Poisson model

## 7.3. Discussion

The time series analysis reported here on the vasculitis population from the Scottish morbidity record shows that there is no seasonal variation in the incidence of AAV in Scotland. These results are complementary to the recent findings by Aiyegbusi et al.,2020 which also reported no seasonal variation in AAV in the Scottish renal biopsy registry (159). Although, the study was smaller (n=339) compared with the routine health dataset used here (n=1622), both results convey the same message. Additionally, findings from chapter 2 highlighted that 50% of studies reporting on the seasonality of AAV observed no seasonal variation. Few of these studies using routine health datasets, like Koldingness et al.,2000 using health records covering 11 hospitals and major regions in Norway reported no seasonal variation in GPA. Chung et al.,2020 also recently reported no seasonal variation in a patient cohort from two major district hospitals in Australia (158). The other half of the literature show mixed results with some reporting higher incidence in winter and spring/summer based on date of symptom onset or diagnosis. Overall, findings from population surveillance data similar to this chapter indicate no clear links between seasonality and AAV.

Interestingly this chapter captured important historic changes in incidence of AAV in Scotland. Specifically, a third of all cases diagnosed with AAV (33%) were detected soon after the introduction of ANCA testing and use in clinical settings by the chapel hill consensus conference committee in 1994 (CHCC,1994) (343). This change in disease classification was captured by a lagging peak in incidence of AAV between 1996 and 2000. A similar peak was seen by Berti et al.,2017 in a twenty-year study from the Minnesota, US (54). Interestingly, the study also reported peaks in incidence of AAV between 2004-2006 and again 2010-2012. Nilsen et al.,2020 also observed oscillating trend in AAV in Norway with peaks in incidence being seen every 2 years (344). In another study from Norway, Koldingness et al.,2000 observed similar peaks in incidence of GPA between 1993 to 1996 using a nationwide routine data (160).

The periodic cycles observed in the SMR01 appear to follow a similar trend from the studies reported above. The analyses in this chapter showed stable minor peaks every 2 years initially (1998-2002) then every 4-5 years between 2002 and 2020. Some of the peaks are partially explained by the noise in the time series (outliers or months with 'higher-than-expected' cases), while others reflect a real trend possibly due to increased awareness of AAV by physicians following an update to vasculitis nomenclature and criteria in 2012 and 2007 (22). Furthermore, the annual incidence in AAV ranged from 5 cases per month to 8 cases per month over the 24-year period.

A major strength of this study is its large sample size, the largest on the subject, compared with other studies assessing the effects of seasonality on AAV. The study covered the whole Scottish population and had one of the longest temporal coverages (1996 to 2020) further capturing the impact of changes in classification criteria on the incidence of AAV in Scotland. Another strength of the results on the periodic cycles of AAV was the use of different modelling techniques to characterise change in disease incidence. These approaches consisted of some of the common time series method aimed at separating a true signal in incidence of AAV from the noise (major outliers). A more robust sensitivity analysis using non-stational cosinor method was used to validate the findings from the loess models approach used here. The benefit of using these approaches was that they allowed for flexibility in capturing the variance in the AAV time series by reporting them remainder values and in a form a rolling confidence interval. Similarly, the results of the seasonal analyses were robustly compared using different modelling approaches (GAM and Poisson regression). Both provided detailed insight into the seasonal pattern of AAV.

Several limitations related to temporal bias and accuracy in ICD codes should be considered when interpreting results of this chapter. Given that the hospital admission data are based on appointments, it is possible that there may be major lags between the time of symptom onset to the date of diagnosis which may not capture the actual time of disease onset. For example, Sreih et al.,2021 recently showed using an open-ended survey (n=375) that most vasculitis patients (73%) report being misdiagnosed in the first few months after the first

onset of symptoms. The study also showed that the median time to diagnosis is 7 months (43). It is possible that the date of diagnosis used to represent the start of the disease may not be a good marker for assessing seasonality in AAV. Additionally, for all the analyses reported here, AAV clinical phenotypes were group into one for pragmatic reason to allow for a power and sample size required for the models that were used. Such grouping did not allow for further insight into possible seasonal differences of AAV based on its clinical phenotypes. For examples, there is data to suggest that there might be a seasonal variation in AAV based according to MPO-ANCA vs PR3-ANCA subtypes. Draibe et al.,2018 showed that the incidence of MPO-ANCA was elevated in winter while no clear seasonal variation was seen for PR3-ANCA (130). The authors suggested that this variation between the clinical phenotypes reflect the heterogeneity in environmental triggers, for instance MPO-ANCA being triggered by seasonal infections (345) while GPA may be a rural disease triggered by environmental exposures associated with rurality (159).

Overall, AAV should be understood as a multifactorial disease with multiple triggers that may vary based on its clinical phenotypes. Investigating seasonal, climatic, and geographic exposures may give insights into its aetiology and further provide insight to its pathogenesis and risk factors.

# Chapter 8: Discussion

## 8.1. Summary of findings

### 8.1.1. Investigating the long-term impact of air pollution on vasculitis and other autoimmune diseases

The overall aim of the review conducted in chapter 2 was to identify studies reporting on the association between outdoor and occupation airborne exposure and impact on vasculitis. A systematic search of studies reporting on these exposures revealed a major gap in evidence reporting on the impact outdoor air pollution and onset of vasculitis. To date, there is only one study conducted on two vasculitis cohort from Nottingham and Norfolk, UK. Findings from this study showed that the effects of air pollution on AAV are variable and are dependent on geography. Specifically, the study observed that AAV patients from Nottingham/Derby were from areas with low air pollution levels, particularly $NO_2$ and $SO_2$ compared with general of the population of that county (odds ratio: 0.84, 95% CI: 0.74 - 0.95 and $NO_2$ odds ratio: 0.85, 95% CI: 0.75 – 0.96). For the Norfolk area, these effects were non-existent as point estimates and confidence intervals were around the null ($PM_{10}$ odds ratio: 0.99, 95%CI: 0.78 - 1.25, $SO_2$: 1.01, 95%CI: 0.86 - 1.20, $NO_2$: 1.02, 95%CI: 0.89 - 1.18).

An important limitation identified from this study was the challenge are the use of appropriate sensitive exposure measures. In this study, the authors used air quality index which groups a range of air pollution concentrations into quintiles. These quintiles are calculated at a group level (lower-super output area) and do not reflect exposure at a postcode level. Additionally, the study had limited geographic coverage, with the use of vasculitis data from two counties in the east-midland, making generalisability difficult. Results from the meta-analysis reported in chapter 2 showed that exposure to occupational airborne pollutants such as crystalline silica, a fine particulate matter with aerodynamic size of 5 microns, was associated with 2-fold increased risk of AAV.

In light of this gap in literature and the evidence about the role of occupation airborne exposure and their effect on AAV, the thesis aimed to investigate the long-term impact of airborne exposure at a population level and their association with vasculitis. Other autoimmune diseases were used for comparative purpose and to see if the effects of airborne exposures are shared across a number of leading inflammatory autoimmune diseases affecting the UK population. In achieving the objective of this thesis, a case-control study design was used with baseline and longitudinal data from UK Biobank (UKB) and Scottish Morbidity Record (SMR01). The thesis author linked longitudinal air pollution data from DEFRA covering a period between 2001 to 2020. Results from UKB revealed that a 1 µg per cubic meter higher average levels of $SO_2$ was associated with 6.4% (95% CI: 4% – 13%) and 3.7% (95% CI: 1.1% – 6.4%) increased odds of vasculitis and rheumatoid arthritis, respectively. This association was not statistically significant after adjusting for multiple comparison, suggesting the need for more validation. Similarly, for systemic lupus erythematosus, 1 µg per cubic meter higher average level in $SO_2$ between 2001 to 2018 was associated with 9.7% (95% CI: - 9.4 – 21.1) increased odds of developing SLE, though this was not statistically significant, in part due to the low case number in UK Biobank cohort.

Chapter 6 provided further validation using routine health data from the Scottish Morbidity Record (SMR01). This data was linked with air pollution data again and focused on two major vasculitis phenotypes, ANCA-associated vasculitis, and Giant cell arteritis. The study showed that $SO_2$ was again significantly associated the onset of vasculitis in Scotland. Specifically, a 1 µg per cubic meter higher average level of $SO_2$ from 2001 to 2020 was associated with 6.86% (odds ratio: 1.0686, 95% CI: 1.007 – 1.134) and 7.6% (odds ratio: 1.076, 95% CI: 1.019 - 1.136) increased odds of vasculitis. This association was again not statistically significant after adjusting for multiple comparisons. $PM_{10}$ and $PM_{2.5}$ were also associated with 6.94% (odds ratio: 1.069, 95% CI: 1.044 – 1.095) and 11% (odds ratio: 1.107, 95% CI: 1.068 – 1.148) increased odds of GCA and this was consistently significant after adjusting for multiple comparison. Despite the overall decrease in air pollution over the past two decades, there was no fall in

the incidence of vasculitis, due its rarity and the need for more time to be able to see significant changes in incidence in vasculitis.

In addition to the analyses on the effects of air pollution on vasculitis, chapter 5 looked at the association between air pollution on blood markers of inflammation. The premise here is that by characterising the effects of air pollution on surrogate markers of inflammation, we may get insight into the pathways involved in the pathogenesis of vasculitis and other autoimmune diseases. Interestingly, findings from these analyses showed that $SO_2$ was significantly associated with blood markers of disease activity. Specifically, 1 μg per cubic meter higher average levels of sulphur dioxide was associated with 3.2% (95% CI: 2.93 – 3.44) higher C-reactive protein (CRP) concentrations and 0.8% (95% CI: 0.79 – 0.96) and 0.9% (95% CI: 0.77 – 0.94) higher total neutrophil and monocyte count. Benzene, $PM_{10}$ and $PM_{2.5}$ were inverse associated with CRP and eosinophil while no association were observed for other air pollutants ($NO_2$ and $NO_x$). It is worth noting, while these estimates were adjusted for potential confounding, they may be affected by unmeasured confounders including time-dependent factors and the overall prevalence morbidities in UK Biobank. Combined, these findings from UK Biobank and SMR01 provide novel results that show that $SO_2$ may be an important risk factor of vasculitis and other autoimmune diseases. Its effect may start in the blood through a mediated change in surrogate markers of inflammation.

### 8.1.2. Explaining geographic variation in the onset of vasculitis and its association with outdoor air pollution

Chapter 5 and 6 presented stratified analyses showing the total effect of air pollution based on UK rural and urban geographies. Results from Chapter 5 showed that the risk associated with $SO_2$ were higher in rural populations for both vasculitis and rheumatoid arthritis. Additionally, among those who reported to spend more than 3 hours outdoors, perhaps due to occupation or physical activity, the effects of $SO_2$ on vasculitis was higher compared with those that reported spending less than 3 hours outdoors. This difference in effects by time spent outdoors were not seen for rheumatoid arthritis or other autoimmune

diseases. Similarly for chapter 6, stratified analyses assessing the association between air pollution and AAV including GCA revealed that the effects of $SO_2$ on AAV was higher in those from rural areas. For both AAV and GCA, the effects of $PM_{10}$ and $PM_{2.5}$ on were elevated in those from rural areas compared with urban areas. Interestingly, the meta-analysis conducted in chapter 2 showed that having worked on a farm for more than 6 months was associated increased risk of AAV. The results reported here are novel and may reflect the distribution of air pollution across the population, with those from rural areas being particularly at risk. Even though majority of participants (>75%) from UK Biobank and the Scotland were from urban areas, vasculitis patients from rural areas were at greater risk for the effects $PM_{10}$, $PM_{2.5}$, and $SO_2$ and onset vasculitis.

### 8.1.3. Explaining seasonal and temporal clusters in ANCA-associated vasculitis occurrence and their association with historic weather data

Chapter 6 presented time series analyses of the temporal variation in the incidence of ANCA-associated vasculitis. Overall, incidence data from the SMR01 showed that incidence of AAV peaked in the early 90's following the introduction of ANCA testing and use in clinical settings. There were two major periodic cycles in AAV, one between 1996-2000 (1$^{st}$ peak) and the second one being in 2017-2018. There were also two troughs, the first in 2004 then again in 2013. When assessing these peaks at shorter time frame (12 months), there were peaks at day 190 and 330 out of the 365 days in a year. When cases were aggregated based on the seasons of the year, there was no significant season variation in AAV (pairwise p(FDR) = 0.66).

## 8.2. Critical appraisal

### 8.2.1. Contribution of this thesis to the literature on the aetiology of vasculitis

This thesis began by summarising existing evidence reporting on the association between environmental exposures and the onset of vasculitis. Most of the studies

reported in chapter 2 were subjective to selection bias and had variable measures of the environment that mainly focused on occupational exposures such as crystalline silica (a fine particulate matter - $PM_5$ - $PM_{10}$), organic and inorganic solvents and dust (186–188,210–213,346,347). These studies used indirect exposure methods through self-reported questionnaire to derive occupation exposures that a person is exposed throughout their life. These studies were often not generalisable in terms of reflecting real world exposure at a population level and were often subject to selection bias (348). This thesis addressed this gap by linking health routine data with environmental surveillance data to investigate the impact of outdoor airborne exposure on vasculitis. With this approach, exposure ascertainment was based on finding quantifiable measures that proximate levels of air pollution at the population level. These exposure records covered the whole population of UK and Scotland, thereby allowing us to design high quality observation studies that is generalisable to the UK population. This thesis is the first and longest linkage study to characterise the impact of air pollution on vasculitis. It makes novel contribution to the literature by providing the first validated results showing the impact of sulphur dioxide and particulate matter on vasculitis (Figure 8.1). It also shows that the effects of air pollution are pronounced in rural areas, smokers and in people who self-report to spend most time outdoors (more than 3 hours a day). Similar effects were also seen with rheumatoid arthritis, thus indicating possible shared mechanism of sulphur dioxide on autoimmune diseases.

The geographic differences seen in air pollution effects on AAV and GCA echoes what others have observed in terms of observing higher incidence of vasculitis in rural areas (98,171,172). The thesis presents generalisable population data that is representative of the rural populations. Although this population made up less than 30% of study participants and patient population, results from this thesis shows the risk attributable to air pollution is greater for these individuals compared with those from urban areas. Evidence from the systematised review also indicated that this population is also at great risk especially those involved in farming and gardening activities. It is possible that this population may be at a great risk for sulphur dioxide and particulate matter due to increased reliance on burning of wood and coal, including use of old machinery for farming with higher

concentration of diesel or other fuel products. There is also evidence from this thesis that air pollution may vary based on deprivation, though this variation is not clear and will require rich spatiotemporal modelling and clustering to understand the impact of air pollution across space and time. Nonetheless, the thesis provides important evidence about who's potentially at risk for AAV due to air pollution. It introduces environmental candidates that can be assessed and investigated to see their role in the triggering onset of vasculitis.

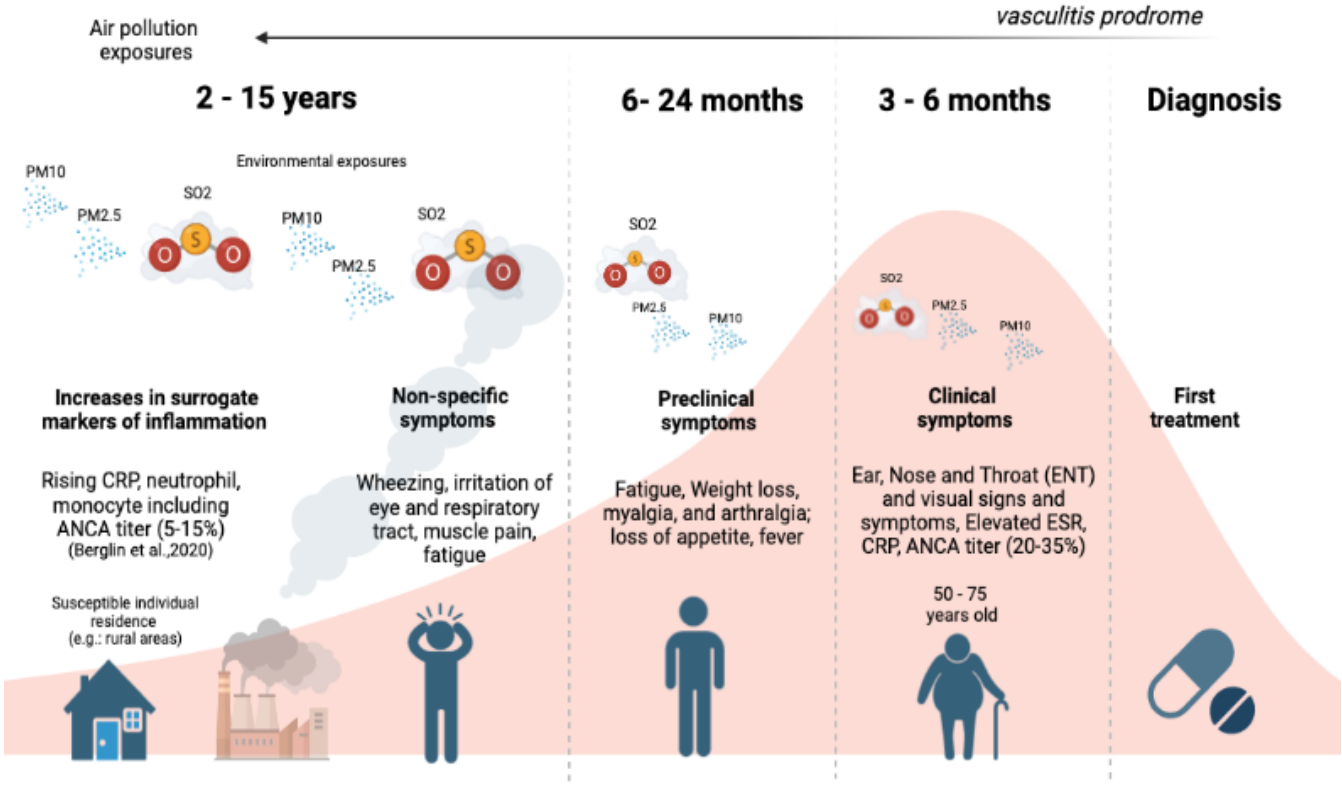



**Figure 8.1:** Schematic overview of the contribution made by this thesis on the role of air pollution and the onset of vasculitis, particularly ANCA-associated vasculitis.

## 8.2.2. Possible mechanism

Evidence from studies reporting on the interaction between genes and environment show that T cell immunity is strongly driven by genetic factors while B cell immunity are mainly influenced by the environment(349). In the context of air pollution, recent studies have shown the first site of immune

response is the lungs which result in increased inflammation and airway damage. Inflammation at the respiratory tract can often lead to systemic inflammation following repeat pollution exposure with age (315,336). Air pollution, as an antigen, can bind Aryl hydrocarbon receptor (AHR) to regulate TH7 and Treg cells, resulting in the production of pro-inflammatory cytokines (IL-1,4,5,6,8,10,13, and TNF alpha) (350). These cytokines stimulate B cell and dendritic cells, resulting in an influx of antibodies and self-reactive T lymphocytes (315).For $SO_2$ specifically, a Brisbane family-system genetic study with 266 pairs of twins (1992 monozygotic and 74 dizygotic pairs) and 165 parents observed a significant association between $SO_2$ and increase expression of the Aryl hydrocarbon receptor (*AHR)* gene (350).In a Canadian biobank study, $SO_2$ was also reported to account for 170 differentially expressed genes enriched in pathways involved in leukocyte migration during chronic inflammation, through a CXCR chemokine activity and G-protein couple receptors. These effects of $SO_2$ remained even after adjusting for ancestral and geographic differences (334). With this, we know CXCR chemokines are important mediators of B cell migration, both peripherally and into the synovium (335). It is possible that $SO_2$-may trigger the onset of vasculitis or other inflammatory autoimmune diseases through cumulative expression of CXCR chemokine activity initiated via the Aryl hydrocarbon receptor or G-protein coupled receptor pathways that leads to chronic inflammation and an antibody mediated response in susceptible individuals.

### 8.2.3. Limitation of air pollution studies in this thesis

The analyses reported in this thesis used DEFRA annual mean pollution data covering the whole UK population. The limitation of using such long aggregate estimates, is the inability to assess the short-term impact of air pollution on disease as well as get insight into seasonal variation in air pollution. In addition, modelled pollution concentrations are imperfect and do not reflect personalised exposure as most people spend most of their time indoors. This assumption can lead to non-differential misclassification of exposure dose and may bias the results towards the null. Stratified analyses from UK biobank provided supported that indeed those who report to spend on average more than three-hours outdoor

were at greater risk, indicating the true effects of outdoor air pollution in driving rare outcomes like vasculitis.

Additionally, it was not possible to access historic postcode data from the SMR01. This meant that the linking of air pollution data was made on the assumption that that people remained at the same address over the chosen study period (2001 to 2020). While this is a limiting factor, data on the UK internal migration suggests that individuals over the age of 40 often remain at one address for between 15 to 21 years (351,352). While this may be true, result from chapter 6 must be interpreted with these limitations. Additionally, the linkage of air pollution data was done based using a four to five-character postcode (postcode sector) with buffer accuracy of 5km radius from the participants address. This was done with the purpose of retaining patient privacy. The limitation is that using a broad exposure range (5km radius) could lead to measurement error and underestimate the true effects of air pollution on vasculitis.

All the results from the statistical models presented in this thesis were from single pollutant models and corrected for multiple comparison. This is common practice in environmental health research as lot of the pollutants are known to be collinear. The results from these models do not allow for combined assessment of the effect of multiple pollutant on the onset of vasculitis. This may be important as the total effects attributable to air pollution may differ between a sum of combined pollution models that recognise simultaneous impact of multiple pollutant on an outcome.

Lastly, the effects of air pollution in UKB study are limited by the fact the outcome data were pooled based on ICD-10 three-character codes which did not allow to assess the varying effects of air pollution on vasculitis clinical phenotype. As seen with the SMR01, the effect of fine particles and gaseous air pollutants were different and varied by marginal degree. It may be true that different air pollutants may affect the AAV phenotypes differently, given their varying mechanism.

### 8.2.4. Limitation of the temporal and seasonal variation study

Like many studies reporting the seasonality and temporal cycles of vasculitis, chapter 7 findings were limited by the information source and variables available from the routine data. The lack of case notes records about the onset of symptoms limited further sensitivity analyses. Previous studies have suggested using date of symptom onset may be a better marker of disease onset. For example, there is data to show there is winter peak in incidence of AAV when using the date of symptom onset (130). This winter peak may be explained by environmental factors like infection and pollution.

Changes in the nomenclature and classification criteria including the coding of AAV in routine health records could have also explain most of the peaks seen in this study. These time-dependent factors limit our ability to infer the role of infectious or environmental related events with the incidence of AAV in Scotland. Unlike Kawasaki disease where the naming and classification has been stable over time (353), previous studies have been able to suggest that some of the major peaks seen for KD in Japan may be attributable to the influenza pandemic, the N1H1 outbreak seen between 2009 and 2010 (354,355). No such inference can be made for AAV.

Furthermore, the seasonal analyses based on AAV clinical phenotype could not be carried due to limited specificity of ICD codes for GPA, MPA and EGPA for clinical meaning. For example, Cook et al., 2022 showed that patients coded as having GPA which is mostly associated with proteinase-3 (PR-3-ANCA) actually were MPA, with over half of them (55%) being MPO-ANCA+ (337). There is increasing recognition that ANCA-subtype may be a better classification of AAV than its clinical phenotype listed in ICD-coding system(356). Unfortunately, routine health data do not provide data on ANCA serotype and therefore it is not possible to assess any seasonal variation based on this criteria.

## 8.3. Clinical and public health implications

A major finding of this thesis is that air pollution and specifically sulphur dioxide and particulate matter may be an important risk factors of vasculitis, especially for people residing in rural areas. Farmers were also seen to be at an increased risk based on existing literature. These findings have implication for air quality management, and health and safety policies as well as clinical guidelines for autoimmune diseases. One such implication is the cost associated with air pollution on healthcare. In 2017, it was estimated that the cost associated with $PM_{2.5}$ to the NHS and social care was £1.68 billion in England (357). This was calculated based on data where there is robust evidence for the effect of $PM_{2.5}$ and cardiorespiratory morbidities and mortality. Additionally, between 2017 to 2025, the total cost associated with air pollution ($PM_{2.5}$) in England, for other diseases where there is less robust evidence, is £2.81 billion (357). For $SO_2$, there are no models to show its implication on healthcare. Findings from this thesis emphasise the need for new cost-benefit analyses to show the impact of $SO_2$ to the NHS, with a special focus on rural health boards where air pollution may account for an increase in the number vasculitis cases including other autoimmune diseases.

Furthermore, previous policies (e.g.: Clean air Act, 1993) have focused on setting air quality standards and targets (264). These targets are currently being met by counties and cities across the UK and have resulted in a 95% reduction of $PM_{2.5}$ and $SO_2$ between 1990 to 2019 (358). While these targets have had many benefits for health and disease prevention (359), new targets are needed given that $PM_{2.5}$, is still one of the leading causes of morbidities and mortality (360,361). A new proposal for revision of the EU ambient air quality directive will aim to lower air quality standards even further to reach the 2050 zero emissions (358,362). This is important because there are disparities in air pollution exposure, with those from residing in deprived areas and certain rural areas being at an increased risk. For new policies to work, they will need to be specific, and should follow low emission zones policies introduced in 2008 (363,364), and with focus on rural areas and more deprived areas.

From a health and safety (H&S) perspective, more robust studies are needed to inform the industrial injuries advisory council (IIAC) guidelines on farming and AAV. IIAC recently looked at the evidence of silica and AAV and concluded that there is no strong evidence to require new guidelines (347). This is mainly due most studies had a low sample size and were subject to selection and publication bias. The council also noted that agricultural workers were a group potentially at risk for higher silica exposure. For example, Lane et al.,2013 reported that 49% of the study subjects exposed to silica were agricultural workers in the UK (213). Hogan et al.,2007 reported that 22% of study subjects were exposed to silica through harvesting crops in the US(211). Given the complex and diverse range of exposures associated with farming, new safety guidelines may be needed to protect farm workers against small particles and other occupation airborne exposure linked with the onset of vasculitis.

From a clinical and research perspective, air pollution needs to be recognised as an important risk factor of autoimmune diseases(315). Recently, members of the European Respiratory Society (ERS) have recognised the importance of including some of the more robust evidence of air pollution in medical textbooks to increase awareness in among new clinicians and help general practitioners identify vulnerable groups such as those that reside near major roads with high traffic intensity or near waste land or near industrial areas (365).

## 8.4. Future directions

More work needs to be done to contribute high quality evidence that show the role of environmental exposures on the onset of vasculitis. Current evidence from nested and non-nested control studies are often limited in power and sample size including generalisability. New studies using population level data should integrate routine health data from tertiary care setting with biobank data, and environmental data covering whole populations. Additionally, linking vasculitis registries with national biobanks using direct linkage methods can help facilitate research into the aetiology of vasculitis. For example, linking the UKIVAS registry with UK Biobank and its rich genetic, biomarker and

environmental data including self-reported exposures data, can allow for more robust research into the pathogenesis of vasculitis. Linkages approaches highlighted in chapter and commonly used in environmental-wide association studies (EWAS) provide necessary methods for aetiology research. To achieve this, major barriers related to the lack of representativeness and accuracy of population data will need to be considered. While routine data offer a representative population coverage with bigger sample sizes; they lack detailed context information pertaining to clinical, social, behavioural and biomarker phenotypes beyond what is captured in a standardise coding system. Similarly, vasculitis registries are great at capturing the necessary clinical phenotypes needed to distinguish the different forms of vasculitis, but they are not always representative and lack self-reported information that can give insight into potential exposures associated with the onset of disease. Registries also lack a population control group needed to achieve an appropriate study design to facilitate an EWAS. Biobanks carry rich longitudinal outcomes data with a broad baseline genetic, biomarker and self-reported data but often are not representative of the general population. Careful planning and implementation will be needed when designing a study addressing aetiology of vasculitis. Future studies should seek linkage expertise so to maximise on the rich health and environmental data resources available for research in the UK and Europe.

# Appendices

# Appendix 1: Medline Search Strategy for the systematised review

Search period 1946 – October 2021

**Keywords:**

[1] SH: Air pollution/or traffic-related pollution/or air pollutants/or atmospheric pollutant/or particulate matter/or nitrogen oxides/or nitrogen dioxide/ or nitrous oxide/ Sulfur Dioxide/ or Sulphur Dioxide/or Carbon Monoxide/ozone/

[2] TW: Outdoor pollut*/or air qual*/ or air contamin∗/or atmospheric pollut∗/or atmosphere contaminat∗/or traffic pollut*/or ambient pollut*/ or traffic-related pollut*/or vehicle emi*
TW: “PM10”/or “PM2.5”/or “NO2”/ or “NOx”/or “SO2/ or “CO”

[3] (1) OR (2)

[4] TW: Temperature/ or Outdoor temperat*/ or ambient temperat*/ or air temperat*/or surface temperat*/or seasonal temperat*/ or season*/ or seasonal*/ or atmosph*/or atmospheric condit*/or weather condit*/ wind/ tropospheric wind

[5] SH: occupational exposures/or occupational dusts/or occupational airborne chemicals/or dust/or silica/or metals/or diesel/or fibres/or asbestos/

[6] (3) OR (4) OR (5)

[7] exp Systemic Vasculitis/or exp Anti-Neutrophil Cytoplasmic Antibody-Associated Vasculitis/or small vessel vasculitis/Medium vessel vasculitis/or large vessel vasculitis/or Autoimmune Vasculitis/ or Vasculitis/ or Vasculitides/ or granulomatosis with polyangiitis/ or microscopic with polyangiitis/ or eosinophilic granulomatosis with polyangiitis/ or Churg Strauss Syndrome/ or Wegener's granulomatosis/ or Giant Cell Arteritis/ or / or Polyarteritis nodosa/Takayasu arteritis

[8] TW: “vasculitis”/or “GPA”/or “MPA”/or “EGPA”/or “WG”/or “GCA”/or “PAN”

[9] Chronic rheumatic disease/or Chronic Inflammatory Rheumatic disease/ or Inflammatory Rheumatic disease

[10] Arthritis, Rheumatoid/

[11] TW: Rheumatoid Arthritis/or Rheumatic Arthritis/

[12] exp Lupus Erythematosus, Systemic/

[13] TW: Systemic Lupus Erythematosus/or Lupus erythematosus/or Lupus Erythematosus Disseminatus

[14] exp Arthritis, Psoriatic/

[15] Psoriatic Arthritis.mp. /or t.w. Psoriatic arthritis
[mp = title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]

[16] Idiopathic Ankylosing Spondylitis.m.p. /or t.w. Ankylosing Spondylitis
[mp = title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]

[17] exp Systemic Vasculitis/or exp Anti-Neutrophil Cytoplasmic Antibody-Associated Vasculitis/or Autoimmune Vasculitis/ or exp Vasculitis

[18] TW: Systemic sclerosis

[19] exp Sclerosis, Systemic/ or exp Scleroderma, Systemic/or exp Rheumatic Diseases/

[20] TW: vasculitis

[21] (4) OR (5) OR (6) OR (7) OR (8) OR (9) OR (10) OR (11) OR (12) OR (13) OR (14) OR (15) OR (16) OR (17) OR (18) OR (19) OR (20)

[22] (6) AND (21)

# Appendix 2: Joanna Briggs Critical Appraisal Checklist for case control studies

The purpose of this appraisal was to assess the methodological quality of the case control studies used to assess the association between air pollution and occupation exposures with vasculitis in chapter 2. JBI helped determine the extent to which a study has addressed the possibility of bias in its design, conduct and analysis.

Reviewer_____Date_____

Author _____Year_____Record Number_____

| | Yes | No | Unclear | Not applicable |
|---|---|---|---|---|
| 1. Were the groups comparable other than the presence of disease in cases or the absence of disease in controls? | ☐ | ☐ | ☐ | ☐ |
| 2. Were cases and controls matched appropriately? | ☐ | ☐ | ☐ | ☐ |
| 3. Were the same criteria used for identification of cases and controls? | ☐ | ☐ | ☐ | ☐ |
| 4. Was exposure measured in a standard, valid and reliable way? | ☐ | ☐ | ☐ | ☐ |
| 5. Was exposure measured in the same way for cases and controls? | ☐ | ☐ | ☐ | ☐ |
| 6. Were confounding factors identified? | ☐ | ☐ | ☐ | ☐ |
| 7. Were strategies to deal with confounding factors stated? | ☐ | ☐ | ☐ | ☐ |
| 8. Were outcomes assessed in a standard, valid and reliable way for cases and controls? | ☐ | ☐ | ☐ | ☐ |
| 9. Was the exposure period of interest long enough to be meaningful? | ☐ | ☐ | ☐ | ☐ |
| 10. Was appropriate statistical analysis used? | ☐ | ☐ | ☐ | ☐ |

Overall appraisal: Include ☐ Exclude ☐ Seek further info ☐

Comments (Including reason for exclusion)

_____

_____

_____

# Appendix 3: Joanna Briggs Critical Appraisal Checklist for analytical cross-sectional studies

Reviewer_____Date_____

Author_____Year_____ Record Number_____

| | Yes | No | Unclear | Not applicable |
|---|---|---|---|---|
| 1. Were the criteria for inclusion in the sample clearly defined? | ☐ | ☐ | ☐ | ☐ |
| 2. Were the study subjects and the setting described in detail? | ☐ | ☐ | ☐ | ☐ |
| 3. Was the exposure measured in a valid and reliable way? | ☐ | ☐ | ☐ | ☐ |
| 4. Were objective, standard criteria used for measurement of the condition? | ☐ | ☐ | ☐ | ☐ |
| 5. Were confounding factors identified? | ☐ | ☐ | ☐ | ☐ |
| 6. Were strategies to deal with confounding factors stated? | ☐ | ☐ | ☐ | ☐ |
| 7. Were the outcomes measured in a valid and reliable way? | ☐ | ☐ | ☐ | ☐ |
| 8. Was appropriate statistical analysis used? | ☐ | ☐ | ☐ | ☐ |

Overall appraisal: Include ☐ Exclude ☐ Seek further info ☐

Comments (Including reason for exclusion)

_____

_____

_____

# Appendix 4: Critical appraisal rating of studies reporting the association between air pollution and silica and AAV and GCA

| | 1. Appropriateness of cases and controls | 2. Cases and controls matched appropriately | 3. Same criteria for identifying cases and controls | 4. Exposure measured in a valid way | 5. Exposure measured in the same way for cases and controls | 6. All major Confounders identified | 7. Confounders adjusted for | 8. Outcomes assessed in standard, valid and reliable way | 9. Exposure period long enough to be meaningful | 10. Appropriate statistical analysis used | 11. Power/Sample size | Grading |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Air pollution** | | | | | | | | | | | | |
| Mahmood-Rao, 2018 | Y | Y | Y | ? | Y | N | N | Y | N | Y | Y | Moderate |
| **Occupation exposures** | | | | | | | | | | | | |
| Lane, 2003 | Y | Y | Y | Y | Y | N | Y | Y | Y | Y | Y | High |
| Stamp, 2015 | Y | Y | Y | Y | Y | N | Y | Y | Y | Y | Y | High |
| Hogan, 2006 | N | Y | Y | N | Y | N | Y | Y | Y | Y | N | Moderate |
| Maritati, 2021 | Y | Y | Y | N | Y | N | N | Y | ? | Y | Y | Moderate |
| Willeke, 2015 | N | Y | Y | Y | Y | ? | ? | Y | ? | Y | N | Moderate |
| Beaudreuil, 2005 | Y | Y | Y | N | Y | ? | N | Y | N | Y | N | Moderate |
| Albert, 2003 | Y | Y | Y | N | Y | N | N | Y | N | Y | N | Moderate |

Y Yes

N No

? Not clear

# Appendix 5: Critical appraisal rating of studies reporting the association between seasonality and AAV and GCA

| | 1.Inclusion criteria clearly defined | 2. Study subjects and settings described clearly | 3. Exposure measured in a valid way | 4. Objective, standard criteria used for measurement of vasculitis | 5. All major Confounders identified | 6. Confounders adjusted for | 7. Outcomes assessed in standard, valid and reliable way | 8. Appropriate statistical analysis used | 9. Power/Sample size | Grading |
|---|---|---|---|---|---|---|---|---|---|---|
| Chung, 2020 | Y | Y | Y | Y | N | N | Y | Y | N | Moderate |
| Aiyegbusi, 2020 | Y | Y | Y | Y | N | N | Y | Y | Y | High |
| González-Gay, 2001 | Y | Y | Y | Y | N | N | Y | Y | N | Moderate |
| Koldingsnes, 2000 | Y | Y | Y | Y | N | N | Y | Y | N | Moderate |
| Richier, 2018 | Y | Y | Y | Y | N | N | Y | Y | N | Moderate |
| González-Gay, 2007 | Y | Y | Y | Y | N | N | Y | Y | Y | High |
| Stamatis, 2019 | Y | Y | Y | Y | N | N | Y | Y | Y | High |
| Khelgi, 2017 | Y | Y | Y | Y | N | N | Y | N | N | Moderate |
| Gokoffski, 2019 | Y | Y | Y | Y | N | N | Y | Y | N | Moderate |
| Konig, 2021 | Y | Y | Y | Y | N | N | Y | Y | Y | High |
| Raheel, 2018 | Y | Y | Y | Y | N | Y | Y | Y | Y | High |
| Mahr, 2006 | N | Y | N | Y | N | N | N | Y | N | Low |
| Draibe, 2018 | Y | Y | Y | Y | N | N | Y | Y | Y | High |
| Frieta-Gilchrist, 2020 | Y | Y | Y | Y | N | Y | Y | Y | Y | High |
| Petursdottir, 1999 | Y | Y | Y | Y | Y | Y | Y | Y | Y | High |

Y Yes

N No

# Appendix 6: Studies reporting the association between air pollution and autoimmune diseases

| Author | Title | Journal |
|---|---|---|
| Bai, 2021 | Acute effects of air pollution on lupus nephritis in patients with systemic lupus erythematosus: A multicentre panel study in China | Environmental Research 195 (2021): 110875 |
| Shin, 2019 | Association between Exposure to Ambient Air Pollution and Rheumatoid Arthritis in Adults | International journal of environmental research and public health 16 (2019) |
| Wu, 2021 | Association between traffic-related air pollution and hospital readmissions for rheumatoid arthritis in Hefei, China: A time-series study | Environmental Pollution 268 (2021): 115628 |
| Park, 2021 | Association of Particulate Matter with Autoimmune Rheumatic Diseases among Adults in South Korea | Rheumatology (Oxford, England) (2021) |
| Zhao, 2019 | Effect of air pollution on hospital admissions for systemic lupus erythematosus in Bengbu, China: a time series study | Lupus 28 (2019): 1541 |
| Hart JE, 2009 | Exposure to traffic pollution and increased risk of rheumatoid arthritis | Environmental health perspectives 117 (2009/07): 1065 |
| Jung, 2019 | Long-term exposure to traffic-related air pollution and systemic lupus erythematosus in Taiwan: A cohort study | Science of the Total Environment 668 (2019): 342 |
| Stojan, 2014 | Outdoor Air Pollution and Systemic Lupus Erythematosus | Arthritis and Rheumatology 72 (2020): 2039 |
| De Roos, 2014 | Proximity to traffic, ambient air pollution, and community noise in relation to incident rheumatoid arthritis | Environmental Health Perspectives 122 (2014): 1075 |

| | | |
|---|---|---|
| Castillo-Ortiz, 2017 | The age of onset of rheumatoid arthritis correlates with air pollution and health expenditure: Results from multinational databases | Annals of the Rheumatic Diseases 76 (2017): 1452 |
| Cakmak, 2021 | The association between air pollution and hospitalization for patients with systemic lupus erythematosus in Chile: A daily time series analysis | Environmental research 192 (2021): 110469 |
| Hart JE, 2011 | The association of ambient air pollution exposures and risk of rheumatoid arthritis: Results from the Swedish EIRA case-control study and the US nurses' health prospective cohort study | Arthritis and Rheumatism 63 (2011) |
| Chen, 2021 | The relationship of polluted air and drinking water sources with the prevalence of systemic lupus erythematosus: a provincial population-based study | Scientific reports 11 (2021): 18591 |
| Jung, 2017 | Air Pollution as a Potential Determinant of Rheumatoid Arthritis: A Population-based Cohort Study in Taiwan | Epidemiology (Cambridge, Mass.) 28 Suppl 1 (2017): S54 |
| Chang, 2016 | Air pollution exposure increases the risk of rheumatoid arthritis: A longitudinal and nationwide study | Environment international 94 (2016): 495 |
| Hart JE, 2013 | Ambient air pollution exposures and risk of rheumatoid arthritis | Arthritis care & research 65 (2013): 1190 |
| Hart JE, 2013 | Ambient air pollution exposures and risk of rheumatoid arthritis: Results from the Swedish EIRA case-control study | Annals of the Rheumatic Diseases 72 (2013): 888 |
| Bernatsky, 2014 | Systemic autoimmune rheumatic disease and PM2.5 air pollution levels in Alberta | Journal of Rheumatology 41 (2014): 1470 |

# Appendix 7: NHS Research Ethics Committee Approval for UK Biobank

NHS
Health Research Authority

**North West - Haydock Research Ethics Committee**
3rd Floor - Barlow House
4 Minshull Street
Manchester
M1 3DZ

Telephone: 02071048103

18 June 2021

Professor Naomi Allen
UK Biobank Limited
Clinical Trial Service Unit and Epidemiological Studies Unit
Nuffield Department of Population Health, The Big Data Institute
University of Oxford, Oxford
OX3 7LF

Dear Professor Allen,

| | |
|---|---|
| **Title of the Research Tissue Bank:** | **UK Biobank: a large scale prospective epidemiological resource** |
| **REC reference:** | **21/NW/0157** |
| **Designated Individual:** | **Mrs Samantha Welsh** |
| **IRAS project ID:** | **299116** |

The Research Ethics Committee reviewed the above application at the meeting held on 08 June 2021. Thank you for attending to discuss the application.

**Ethical opinion**

The members of the Committee present gave a favourable ethical opinion of the above research tissue bank on the basis described in the application form and supporting documentation, subject to the conditions specified below.

This application was for the renewal of a Research Tissue Bank application. The previous REC Reference number for this application was **16/NW/0274.**

**Conditions of the favourable opinion**

The favourable opinion is subject to the following conditions being met prior to the start of the Research Tissue Bank.

# Appendix 8: NHS Scotland Public Benefit and Privacy Panel for Health and Social Care Committee Amendment Approval for VOICES Study to link environment data

**Public Benefit and Privacy Panel for Health and Social Care**
phs.pbpp@phs.scot
www.informationgovernance.scot.nhs.uk

Dr Rosemary Hollick
Aberdeen Centre for Health Data Science
Rm 107 Health Sciences Building
School of Medicine, Medical Science and Nutrition
University of Aberdeen
Foresterhill
Aberdeen
AB25 2ZD

Date: 9th May 2022
Ref: 1819-0069

Dear Dr Hollick,

Re Application: Effective healthcare delivery in rare rheumatic disease: evaluating models of care for systemic vasculitis
Version: v9

Further to your approval issued by the Public Benefit and Privacy Panel for Health and Social Care (HSC-PBPP) on 14th January 2020, I am writing to confirm that we accept the amendment(s) to the proposal received by the HSC-PBPP on 19th April 2022.

The approved amendments are:

- Add postcode sector as a variable for the SMR01
- Add ECOSS variables.
- Add pollution and weather data.

Please note that any conditions attached to your original approval remain in place and you should continue to comply with those conditions outlined in the approval letter. It is the responsibility of the applicant and their organisation to ensure that their study always complies with current legislation during the study.

This approval is given to process data, as specified in the approved application form version specified in this letter, until 1st February 2024 and is limited to this.

Requests for access to NHS Scotland data as part of this approved application should be supported by providing a copy of your approval letter and approved application to the relevant local board contacts/data providers.

I would take this opportunity to remind you of the declaration you have made in your application committing you to undertakings in respect of information governance, confidentiality and data protection.

Yours sincerely,

Phil Dalgleish
Depute Panel Manager
NHS Scotland Public Benefit and Privacy Panel for Health and Social Care
Email: phs.pbpp@phs.scot

Cc: Professor Sildaditya Bhattacharya, Main contact for Lead Organisation

# Appendix 9: NHS Scotland Public Benefit and Privacy Panel for Health and Social Care Committee Original Approval for VOICES Study

**Public Benefit and Privacy Panel for Health and Social Care**
nss.PBPP@nhs.net
www.informationgovernance.scot.nhs.uk

NHS SCOTLAND

Dr Rosemary Hollick
Aberdeen Centre for Health Data Science
Rm 107 Health Sciences Building
School of Medicine, Medical Science and Nutrition
University of Aberdeen
Foresterhill
Aberdeen
AB25 2ZD

Date: 14th January 2020
Ref: 1819-0069

Dear Dr Hollick,

**Re Application: Effective healthcare delivery in rare rheumatic disease: evaluating models of care for systemic vasculitis**
**Version: v2**

Thank you for your application for consideration by the Public Benefit and Privacy Panel for Health and Social Care. Your application has undergone proportionate governance review and has been approved.

This approval is given to process data as specified in the approved version of the application, and is limited to this. Approval is valid for the period specified in your application until 28th February 2022. You are required to notify the Panel Manager, via your research coordinator, of any proposed changes to your proposal, e.g. purpose or method of processing, data or data variables being processed, study cohorts, individuals accessing and processing data, timescales, technology/infrastructure.

On conclusion of your proposal, as part of NHS Scotland Governance and monitoring we will require you to complete an End of Project reporting form to demonstrate that you have complied with the obligations outlined e.g. data destruction or submission of references for publications of findings.

I would take this opportunity to remind you of the declaration you have made in your application form committing you to undertakings in respect of information governance, confidentiality and data protection. It is the responsibility of the applicant and their organisation to ensure that their study complies with current legislation at all times during the study.

**Requests for access to NHS Scotland data as part of this approved application must be supported by providing a copy of your approval letter and approved application to the relevant local board contacts and/or data providers.**

Please note that summary information about your application and its approval, including the title and nature of your proposal, will be published on the panel website (www.informationgovernance.scot.nhs.uk).

I hope that your proposal progresses well.

Yours sincerely,

Dr Marian Aldhous

Panel Manager
NHS Scotland Public Benefit and Privacy Panel for Health and Social Care
Email: nss.PBPP@nhs.net

Cc: Professor Sildaditya Bhattacharya, Main contact for Lead Organisation

# Appendix 10: List of data sources used in this thesis.

| Dataset/source Name | Data Controller (Organisation) |
|---|---|
| SMR01 | Public health Scotland https://www.ndc.scot.nhs.uk/Data-Dictionary/SMR-Datasets/SMR01-General-Acute-Inpatient-and-Day-Case/ |
| CHI vital status register | Public health Scotland https://www.ndc.scot.nhs.uk/Data-Dictionary/SMR-Datasets//Patient-Identification-and-Demographic-Information/Community-Health-Index-Number/ |
| SIMD | Public health Scotland and Office of national statistics (ONS) https://www.isdscotland.org/products-and-services/gpd-support/deprivation/simd/ |
| UK Deprivation | ONS and UK Data Service census support http://geoconvert.ukdataservice.ac.uk/ |
| URBAN/RURAL | Public health Scotland<br><br>ONS http://geoconvert.ukdataservice.ac.uk/ |
| Air pollution | UK department for Environment, Food and Rural Affairs (DEFRA) https://uk-air.defra.gov.uk/data/pcm-data |
| Weather | Public Weather Data (National data centre for atmospheric and earth observation research (CEDA) and UK Met office (Metoffice)) https://www.metoffice.gov.uk/research/climate/maps-and-data/data/haduk-grid/datasets |

| | |
|---|---|
| | https://catalogue.ceda.ac.uk/uuid/4dc8450d889a491ebb20e724debe2dfb |

# Appendix 11: List of vasculitis phenotypes used to define systemic vasculitis and other autoimmune disease that served as disease comparators in Chapter 5 from UK Biobank

| UK Biobank field | Data field code | Description | ICD 3-Character description | 4-Character description | Case definition |
|---|---|---|---|---|---|
| 2413 | 131890 | Date ICDM30 first reported | Polyarteritis nodosa and related conditions (M30) | M30.0 Polyarteritis nodosa | Systemic Vasculitis |
| | | | | M30.1 Polyarteritis with lung involvement (Churg-Strauss) | |
| | | | | M30.2 Juvenile Polyarteritis | |
| | | | | M30.3 Mucocutaneous lymph node syndrome | |
| | | | | M30.8 other conditions related to polyarteritis nodosa | |
| | 131892 | Date ICDM31 first reported | Other necrotizing vasculopathies (M31) | M31.0 Hypersensitivity angiitis (Goodpasture syndrome) | |
| | | | | M31.1 Thrombotic microangiopathy | |
| | | | | M31.3 Wegener granulomatosis | |
| | | | | M31.4 Aortic arch syndrome (Takayasu) | |
| | | | | M31.5 Giant cell arteritis with polymyalgia rheumatica | |
| | | | | M31.6 Other giant cell arteritis | |
| | | | | M31.7 Microscopic polyangiitis | |
| | | | | M31.8 Other specified necrotizing vasculopathies | |
| | | | | M31.9 Necrotizing vasculopathy, unspecified | |
| | 131848 | Date ICDM05 first reported | Seropositive rheumatoid arthritis (M05) | M05.0 Felt syndrome | Rheumatoid Arthritis |
| | | | | M05.1 Rheumatoid lung disease | |
| | | | | M05.2 Rheumatoid vasculitis | |

| | | | | | |
|---|---|---|---|---|---|
| | | | | M05.3 Rheumatoid arthritis with involvement of other organs and systems | |
| | | | | M05.8 Other seropositive rheumatoid arthritis | |
| | | | | M05.9 Seropositive rheumatoid arthritis, unspecified | |
| | | | | | |
| | 131850 | Date ICDM06 first reported | Other rheumatoid arthritis (M06) | M06.0 Seronegative rheumatoid arthritis | |
| | | | | M06.1 Adult-onset Still disease | |
| | | | | M06.2 Rheumatoid bursitis | |
| | | | | M06.3 Rheumatoid nodule | |
| | | | | M06.4 Inflammatory polyarthropathy | |
| | | | | M06.8 Other specified rheumatoid arthritis | |
| | | | | M06.9 Rheumatoid arthritis, unspecified | |
| | 131894 | Date ICDM32 first reported | Systemic Lupus Erythematosus (M32) | M32.0 Drug-induced systemic lupus erythematosus | Systemic Lupus Erythematosus |
| | | | | M32.1 Systemic lupus erythematosus with organ or system involvement | |
| | | | | M32.8 Other forms of systemic lupus erythematosus | |
| | | | | M32.9 Systemic lupus erythematosus, unspecified | |
| | 131913 | Date ICDM45 first reported | Ankylosing Spondylitis (M45) | M45 Ankylosing Spondylitis<br>*Incl:* Rheumatoid arthritis of spine | Ankylosing Spondylitis |
| | 131852 | Date ICDM07 first reported | Psoriatic and enteropathic arthropathies (M07) | M07.0 Distal interphalangeal psoriatic arthropathy | Psoriatic Arthritis |
| | | | | M07.1 Arthritis mutilans | |
| | | | | M07.2 Psoriatic spondylitis | |
| | | | | M07.3 Other psoriatic spondylitis | |
| | | | | M07.4 Arthropathy in Crohn disease | |
| | | | | M07.5 Arthropathy in ulcerative colitis | |

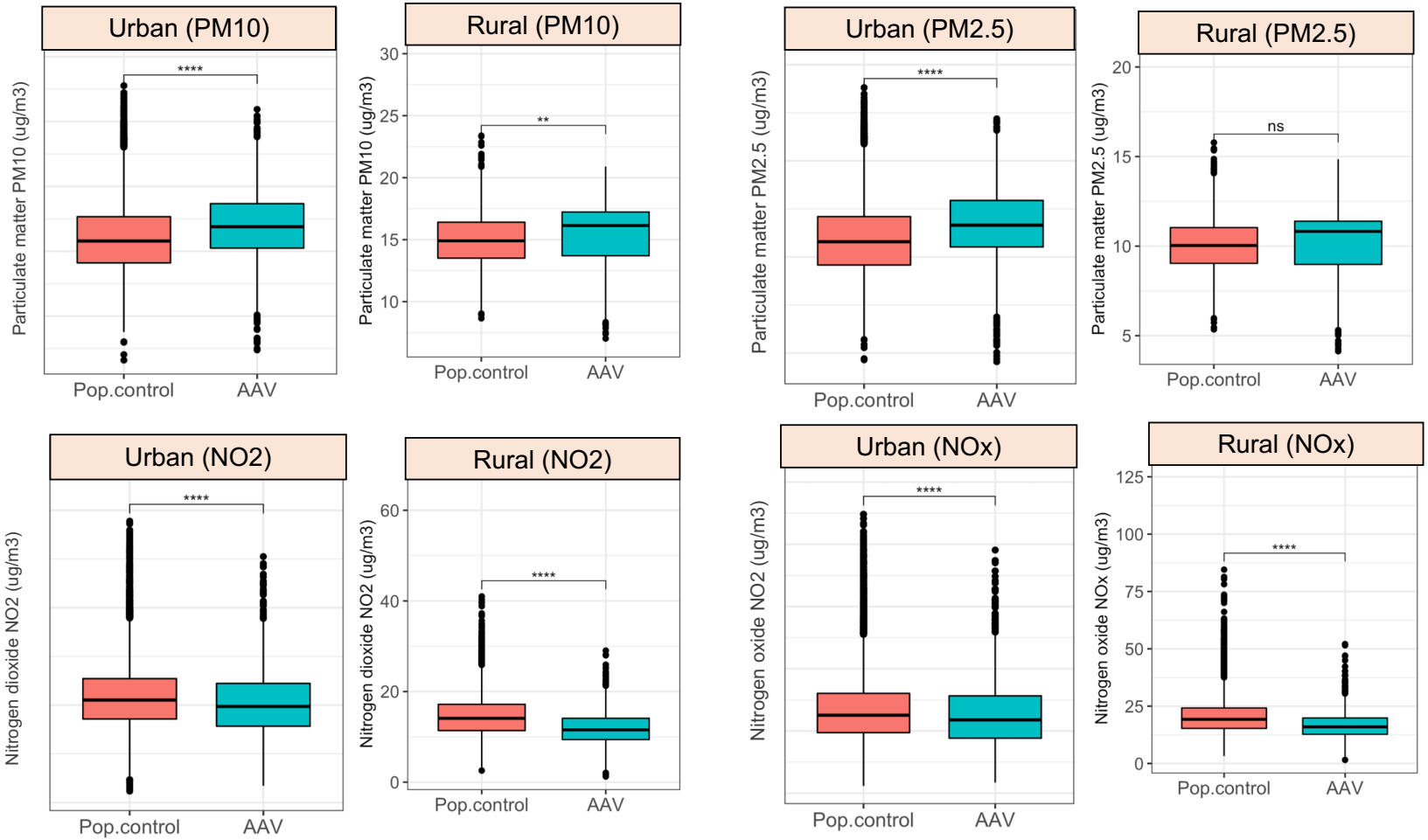
| | | | | | |
|---|---|---|---|---|---|
| | | | | M07.6 Other enteropathic arthropathies | |

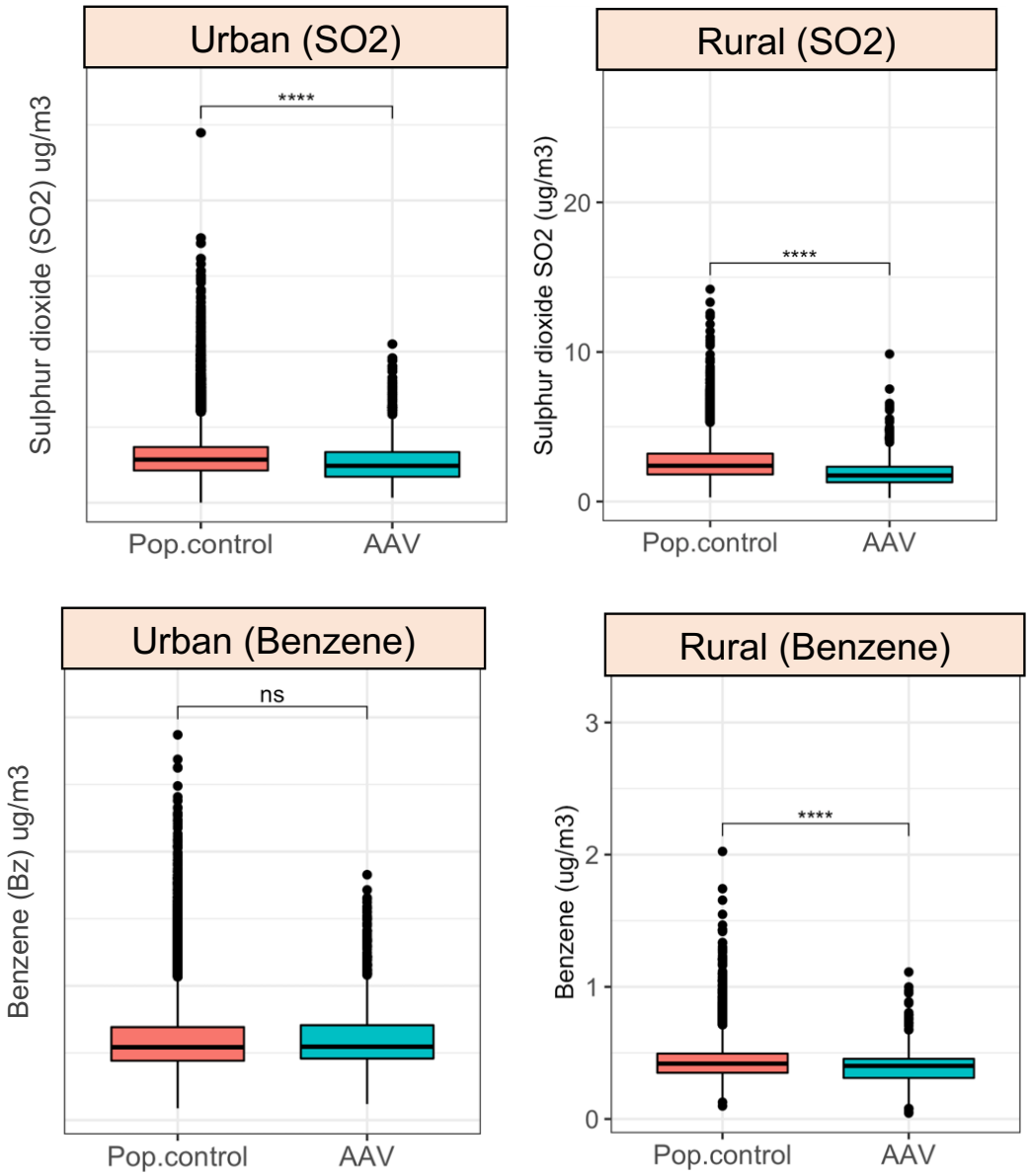
# Appendix 12: List of non-inflammatory autoimmune diseases from UK Biobank used as controls in Chapter 4 pilot study

| UK Biobank field | Data field code | Description | ICD 3-Character description | ICD 4-Character description | Clinical Control definition for this study |
|---|---|---|---|---|---|
| 2413 | 131868 | ICDM15 first reported | Polyarthrosis (M15) | M15.0 Primary generalised (osteo) arthrosis | Polyarthritis |
| | | | | M15.1 Heberden nodes (with arthropathy) | |
| | | | | M15.2 Bouchard nodes (with arthropathy) | |
| | | | | M15.3 Secondary multiple arthrosis | |
| | | | | M15.4 Erosive (osteo) arthrosis | |
| | | | | M15.8 Other polyarthrosis | |
| | | | | M15.9 Polyarthrosis, unspecified | |
| | 131870 | ICDM16 first reported | Coxarthrosis (arthrosis of hip) (M16) | M16.0 Primary coxarthrosis, bilateral | Osteoarthritis |
| | | | | M16.1 Other primary coxarthrosis | |
| | | | | M16.2 Coxarthrosis resulting from dysplasia, bilateral | |
| | | | | M16.3 Other dysplastic coxarthrosis | |
| | | | | M16.4 Post-traumatic coxarthrosis, bilateral | |
| | | | | M16.5 Other post-traumatic coxarthrosis | |
| | | | | M16.6 Other secondary coxarthrosis, bilateral | |
| | | | | M16.7 Other secondary coxarthrosis | |
| | | | | M16.9 Coxarthrosis, unspecified | |
| | 131872 | ICDM17 first reported | Gonarthrosis (arthrosis of knee) (M17) | M17.0 Primary gonarthrosis, bilateral | |
| | | | | M17.1 Other primary gonarthrosis | |
| | | | | M17.2 Post-traumatic gonarthrosis, bilateral | |
| | | | | M17.3 Other post-traumatic gonarthrosis | |
| | | | | M17.4 Other secondary gonarthrosis, bilateral | |
| | | | | M17.5 Other secondary gonarthrosis | |
| | | | | M17.9 Gonarthrosis, unspecified | |

| | | | | | |
|---|---|---|---|---|---|
| | 131874 | ICDM18 first reported | Arthrosis of first carpometacarpal Joint (M18) | M18.0 Primary arthrosis of first carpometacarpal joints, bilateral | |
| | | | | M18.1Other primary arthrosis of first carpometacarpal joint | |
| | | | | M18.2 Post-traumatic arthrosis of first carpometacarpal joints, bilateral | |
| | | | | M18.3 Other post-traumatic arthrosis of first carpometacarpal | |
| | | | | M18.4 Other secondary arthrosis of first carpometacarpal joints, bilateral | |
| | | | | M18.5 Other secondary arthrosis of first carpometacarpal joint | |
| | | | | M18.9 Arthrosis of first carpometacarpal joint, unspecified | |
| | 131876 | ICDM19 first reported | Other arthrosis | M19.0 Primary arthrosis of other joints | |
| | | | | M19.1 Post-traumatic arthrosis of other joints | |
| | | | | M19.2 Other secondary arthrosis | |
| | | | | M19.8 Other specified arthrosis | |
| | | | | M19.9 Arthrosis, unspecified | |

# Appendix 13: Characteristics of air pollution exposure by urban and rural classification of AAV cases from UKIVAS and UKB controls in Chapter 4




Urban (SO2)
Sulphur dioxide (SO2) ug/m3
****
Pop.control
AAV
Rural (SO2)
Sulphur dioxide SO2 (ug/m3)
20
10
0
****
Pop.control
AAV
Urban (Benzene)
Benzene (Bz) ug/m3
ns
Pop.control
AAV
Rural (Benzene)
Benzene (ug/m3)
3
2
1
0
****
Pop.control
AAV

# Appendix 14: List of study variables selected from the UK Biobank and their measurement.

| UK Biobank field | Variable | Description | Variable format |
|---|---|---|---|
| **Socio-demographic** | | | |
| 21022 | Age at recruitment | Derived variable (dd/mm/yyyy) based on the date of birth and the date of attending initial assessment centre | Continuous<br>(37-73 years old) |
| 31 | Sex | Acquired from the central registry at recruitment, but in some cases updated by the participant. Hence this field may contain a mixture of the sex the NHS had recorded for the participant and self-reported sex. | Binary<br>Male<br>Female |
| 189 | Townsend deprivation index | Townsend deprivation index was calculated immediately prior to participant joining UK Biobank. Based on the preceding national census output areas. Each participant is assigned a score corresponding to the output area in which their postcode is located | Ordinal (continuous transformed to quintiles)<br>1- Least deprived<br>5- Most deprived |
| 20116 | Smoking status | Self-reported smoking status | Categorical<br>Never<br>Previous<br>Current |
| 6138 | Education | Self-reported qualification - touchscreen question "Which of the following | Binary<br>Degree or |

| | | | |
|---|---|---|---|
| | | qualifications do you have? (You can select more than one)" – College/university degree, A level or equivalent, O levels/GCSE or equivalent, CSEs or equivalent, NVQ or equivalent, Other professional qual (nursing/teaching), None of the above, prefer not to answer) | No degree |
| 874<br>864<br>894<br>884<br>914<br>904 | Physical activity | Derived variable generated from self-reported factors from the International Physical Activity Questionnaire short form. This included different physical activity type and duration (such as walking, moderate and vigorous physical activity, strenuous sports…etc). These factors were converted into a single measure of total physical activity in metabolic equivalent of task (MET) – hours per week weighted by intensity (walking, moderate or vigorous) | Continuous<br>(hours/week) |
| 1050<br>1060 | Time outdoors | Derived variable based on the average time (in hours) spent outdoors in both winter and summer<br>Touchscreen question "In a typical DAY in summer, how many hours do you spend outdoors?" – The same question was asked for winter.<br><br>If the participant activated the Help button, they were shown the message: *If the time you spend outdoors in summer varies a lot, give the average time per day. For example, if you spend 1 hour a day on each weekday and 4 hours a day on the* | Continuous<br>1hour – 24hrs<br><br>If answer < 0 then rejected<br><br>If answer > 24 then rejected<br><br>If answer > 10 then participant asked to confirm |

| | | | |
|---|---|---|---|
| | | *weekend, the total hours in a week is 13 (5 + 8), so you spend approximately 2 hours a day*<br><br>Coding 100329 defines 3 special values:<br><br>• -10 represents "Less than an hour a day"<br>• -1 represents "Do not know"<br>• -3 represents "Prefer not to answer" | |
| 1558 | Alcohol intake status | Self-reported alcohol drinking frequency status (per week, month or occasionally)<br><br>Touchscreen question "About how often do you drink alcohol?" If the participant activated the Help button, they were shown the message:<br>*If this varies a lot, please provide an average considering your intake over the last year.* | Categorical<br>Daily or almost daily<br>3-4 times a week<br>1-2 times a week<br>1-3 times a month<br>Special Occasion only<br>Never<br>Prefer not to answer (*treated as missing data*) |
| 21000 | Ethnicity | Self-reported ethnicity | Categorical<br>White<br>Mixed<br>Asian or Asian British<br>Chinese<br>Black or Black British<br>Other ethnic group<br>Do not know<br>Prefer not to answer *(treated as missing data)* |
| **Environmental measures** | | | |

| | | | |
|---|---|---|---|
| 20118 | Home area population density | The classification is derived by combining each participants home postcode with data generated from the 2001 census from the Office of National Statistics, using the Geoconvert tool from Census Dissemination Unit. | Categorical<br>England, Wales and Scotland (Urban, Town and Fringe, Village, Hamlet and isolated dwelling, rural mix) |
| 24014 | Close to major road | Indicator variable indicating whether a residential address is within 50 metres of a class 1 or 2 type road and/or within 100 metres of a class 0 road, based upon central road network. The Central road network is taken from Eurostreets version 3.1 digital road network (scale 1:10000), derived from the TeleAtlas MultiNet TM dataset for the year 2008. All roads of class 0, 1 and 2 (motorways, main roads of major importance and other main roads) were classified as major roads. Based upon local knowledge classes 3 and 4 (secondary roads and local connecting roads) were also classified as major roads. | Binary<br>Yes<br>No |
| 24012 | Inverse distance to the nearest major road (2008) | Inverse distance to the nearest major road was generated based on a local road network (1/m). The definition of a major road for the local road network is a road with traffic intensity greater than 5000 motor vehicles per 24 hours. The local road network is taken from the Ordnance Survey Meridian 2 road network (scale 1:50000, 1 metre accuracy), 2009. | Continuous<br>(0.000 – 1.000) |
| 24013<br>24011<br>24009 | Total Traffic intensity, | Traffic intensity is the average total number of motor vehicles per 24 hours on the nearest major road based upon a local | Continuous<br>(motor vehicles per 24 hours) |

| | | | |
|---|---|---|---|
| | Traffic intensity on the nearest major road,<br>Traffic intensity on the nearest road | road network. Traffic count data is from the Road Traffic Statistics Branch at the Department for Transport attached to the local road network. Traffic data for unmonitored links were estimates based on surrounding monitored links. The local road network is taken from the Ordnance Survey Meridian 2 road network (scale 1:50000, 1 metre accuracy), 2009. The definition of a major road for the local road network is a road with traffic intensity greater than 5000 motor vehicles per 24 hours. | |

# Appendix 15: Incidence rate of vasculitis in UK Biobank

| | | | | 95% Confidence interval | |
|---|---|---|---|---|---|
| | | n | Incidence rate per 100,000 person-years | Lower | Upper |
| **Gender** | | | | | |
| | Male | 528 | 12.264 | 10.895 | 13.806 |
| | Female | 274 | 20.005 | 18.369 | 21.786 |
| **Degree** | | | | | |
| | Yes | 585 | 12.434 | 10.806 | 14.307 |
| | No | 195 | 18.294 | 16.870 | 19.838 |
| **Ever smoker** | | | | | |
| | Yes | 395 | 18.344 | 16.634 | 20.231 |
| | No | 401 | 14.851 | 13.456 | 16.390 |

# Appendix 16: Comparison of air pollution exposures between cases and controls in UK Biobank

| | Systemic Vasculitis *n* = 1,206 | Rheumatoid Arthritis *n* = 6,451 | Systemic Lupus Erythematosus *n* = 401 | Ankylosing Spondylitis *n* = 624 | Psoriatic Arthritis n=812 | Population Controls *n* = 487,942 | P value |
|---|---|---|---|---|---|---|---|
| PM10 | 15.27 (13.62, 16.98) | 15.36 (13.80, 16.93) | 15.44 (13.97, 17.28) | 15.49 (13.98, 17.03) | 15.00 (13.57, 16.62) | 15.50 (13.81, 17.07) | <0.001 |
| PM2.5 | 10.31 (9.16, 11.61) | 10.40 (9.26, 11.59) | 10.52 (9.48, 11.84) | 10.44 (9.37, 11.59) | 10.16 (9.06, 11.31) | 10.52 (9.29, 11.74) | <0.001 |
| NO2 | 19.23 (15.30, 23.11) | 19.88 (16.13, 23.57) | 20.30 (16.61, 24.27) | 19.96 (16.56, 23.71) | 19.42 (15.89, 22.78) | 19.65 (15.76, 23.49) | <0.001 |
| NOx | 28.73 (21.71, 36.22) | 29.89 (23.03, 37.05) | 31.08 (23.93, 38.28) | 30.25 (23.82, 37.54) | 29.22 (22.64, 35.82) | 29.48 (22.39, 37.03) | <0.001 |
| SO2 | 3.83 (3.13, 4.64) | 3.94 (3.22, 4.69) | 4.02 (3.22, 4.76) | 4.03 (3.31, 4.80) | 3.82 (3.15, 4.66) | 3.81 (3.10, 4.60) | <0.001 |

# Appendix 17: Baseline characteristics of participants with a vasculitis or other autoimmune disease occurring after enrolment into UK Biobank. These participants were included in the time to event (cox regression) results reported in chapter 5

| | **Systemic Vasculitis** ***n*** **=** 802 | **Rheumatoid Arthritis** ***n*** **=** 3,588 | **Systemic Lupus Erythematosus** ***n*** **=** 158 | **Ankylosing Spondylitis** ***n*** **=** 294 | **Psoriatic Arthritis** **n=**571 | **Population Controls** ***n*** **=** 487,914 |
|---|---|---|---|---|---|---|
| **Age** mean (±SD), years | 61.78 (6.06) | 59.27 (7.35) | 57.25 (7.93) | 58.18 (8.01) | 56.67 (7.76) | 56.46 (8.11) |
| **Gender** female, *n* (%) | 274 (34.2) | 1155 (32.2) | 27 (17.1) | 152 (51.7) | 247 (43.3) | 223909 (45.9) |
| **Deceased,** n (%) | 87 (10.85) | 296 (8.25) | 19 (12.03) | 25 (8.50) | 30 (5.25) | 19,350 (3.97) |
| **Follow up (in years)** | | | | | | |
| Median (p25, p75) | 4.89 (2.951, 6.71) | 5.09 (3.052, 6.687) | 4.73 (2.997, 6.819) | 4.97 (2.609, 6.704) | 4.94 (2.343, 6.712) | 9.92 (9.232, 10.606) |
| **Ethnicity,** *n* (%) | | | | | | |
| White | 759 (95.1) | 3337 (93.8) | 142 (90.4) | 277 (95.2) | 549 (96.8) | 459009 (94.6) |
| Asian or Asian British | 15 (1.9) | 91 (2.6) | 4 (2.5) | 3 (1.0) | 7 (1.2) | 9555 (2.0) |
| Black or Black British | 9 (1.1) | 73 (2.1) | 9 (5.7) | 4 (1.4) | 1 (0.2) | 7831 (1.6) |
| Chinese | 3 (0.4) | 10 (0.3) | 0 (0.0) | 0 (0.0) | 3 (0.5) | 1541 (0.3) |
| Mixed or other ethnic groups | 12 (1.5) | 47 (1.3) | 2 (1.3) | 7 (2.4) | 7 (1.2) | 7301 (1.5) |
| **Educated to degree level** | | | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| Yes (%) | 195 (25.0) | 774 (22.2) | 39 (25.7) | 64 (22.6) | 148 (26.2) | 157762 (33.0) |
| **Townsend score quintile** | | | | | | |
| 1 Least deprived, *n* (%) | 172 (21.5) | 603 (16.8) | 22 (13.9) | 38 (12.9) | 105 (18.5) | 98100 (20.1) |
| 5 Most deprived | 169 (21.2) | 903 (25.2) | 48 (30.4) | 91 (31.0) | 131 (23.0) | 96705 (19.8) |
| **Total Physical activity (METhrs/week)** | | | | | | |
| Median (p25, p75) | 23.97 (11.25, 51.10) | 25.90 (10.55, 58.54) | 27.55 (10.90, 47.55) | 32.17 (12.52, 66.65) | 20.29 (9.60, 51.02) | 28.00 (12.60, 57.70) |
| **Smoking status** | | | | | | |
| Never, n (%) | 395 (49.6) | 1623 (45.7) | 73 (46.5) | 132 (45.5) | 259 (45.5) | 266687 (55.0) |
| Previous | 297 (37.3) | 1424 (40.1) | 62 (39.5) | 121 (41.7) | 243 (42.7) | 167301 (34.5) |
| Current | 104 (13.1) | 507 (14.3) | 22 (14.0) | 37 (12.8) | 67 (11.8) | 51104 (10.5) |
| **Alcohol Frequency** | | | | | | |
| Never, n (%) | 48 (6.0) | 230 (6.4) | 12 (7.7) | 22 (7.5) | 33 (5.8) | 21432 (4.4) |
| Previous | 37 (4.6) | 236 (6.6) | 8 (5.2) | 20 (6.8) | 46 (8.1) | 17085 (3.5) |
| Current | 714 (89.4) | 3102 (86.9) | 135 (87.1) | 250 (85.6) | 490 (86.1) | 447797 (92.1) |
| **Time Spent Outdoors (in hrs)** | | | | | | |
| Median (p25, p75) | 2.50 (2.00, 4.00) | 2.50 (1.50, 4.00) | 2.50 (1.50, 3.50) | 2.50 (1.50, 4.50) | 2.50 (1.50, 4.00) | 2.50 (1.50, 3.75) |
| **Close to major road (2008)** | | | | | | |
| Median (p25, p75) | 45 (5.7) | 285 (8.1) | 13 (8.3) | 24 (8.3) | 40 (7.1) | 34303 (7.1) |
| **Traffic intensity on nearest major road (2008) (vehicles/24 h)** | | | | | | |
| Median (p25, p75) | 17,580 (13,050,26,381) | 17,074 (12,750,25,112) | 17,584 (12,738, 27,267) | 17,738 (13,015,25,189) | 17,677.00 (12,1925,25,828) | 17,103.00 (12,591,25,260) |
| **PM10 ($_{2001-2018}$) (µg/m3)** | | | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| Median (p25, p75) **PM2.5 (2002-2018) (µg/m$^3$)** | 15.21 (13.59, 16.85) | 15.37 (13.84,16.90) | 15.46 (14.14, 17.30) | 15.27 (13.88, 16.96) | 15.21 (13.60, 16.62) | 15.50 (13.81, 17.07) |
| Median (p25, p75) **NO2 (2002-2018) (µg/m$^3$)** | 10.29 (9.11, 11.50) | 10.41 (9.28, 11.55) | 10.51 (9.53, 11.82) | 10.40 (9.33, 11.49) | 10.30 (9.07, 11.29) | 10.52 (9.29, 1.74) |
| Median (p25, p75) **NOx (2002-2018) (µg/m$^3$)** | 19.26 (15.30, 22.87) | 19.88 (16.19, 23.54) | 20.98 (16.89, 23.63) | 19.86 (16.26, 23.90) | 19.43 (15.88, 22.53) | 19.65 (15.76, 23.49) |
| Median (p25, p75) **SO2 (2002-2018) (µg/m$^3$)** | 28.69 (21.71, 35.60) | 29.91 (23.14, 36.96) | 32.01 (24.21, 37.54) | 29.83 (23.27, 37.85) | 29.11 (22.63, 35.38) | 29.48 (22.39, 37.03) |
| Median (p25, p75) **Benzene (2003-2018) (µg/m$^3$)** | 3.82 (3.13, 4.62) | 3.95 (3.23, 4.67) | 4.16 (3.44, 4.73) | 4.05 (3.35, 4.75) | 3.80 (3.14, 4.66) | 3.81 (3.10, 4.60) |
| Median (p25, p75) | 1.60 (1.27, 1.96) | 1.65 (1.31, 2.02) | 1.74 (1.40, 2.12) | 1.66 (1.31, 2.04) | 1.57 (1.26, 1.92) | 1.62 (1.27, 2.02) |

# References


1. Rodriguez-Pla A, Rossello-Urgell J. Trend and geographic disparities in the mortality rates of primary systemic vasculitis in the united states from 1999 to 2019: A population-based study. J Clin Med. 2021;10(8):1–13.
2. Taimen K, Mustonen A, Pirilä L. The Delay and Costs of Diagnosing Systemic Vasculitis in a Tertiary-Level Clinic. Rheumatol Ther. 2021;8(1):233–42.
3. Willan R. Purpura In: On cutaneous diseases. Kimber & Conrad. 1808;1:452–71.
4. Nunnelee JD. Henoch-Schönlein purpura: a review of the literature. Clin Excell Nurse Pract. 2000 Mar;4(2):72—75.
5. Iglesias-Gamarra A, Penaranda E, Espinoza LR. Vasculitides throughout history and their clinical treatment today. Curr Rheumatol Rep. 2011;13(6):465–72.
6. De Virgilio A, Greco A, Magliulo G, Gallo A, Ruoppolo G, Conte M, et al. Polyarteritis nodosa: A contemporary overview. Autoimmun Rev. 2016;15(6):564–70.
7. Matteson EL, Kluge FJ. Think clearly, be sincere, act calmly: Adolf Kussmaul (February 22, 1822–May 28, 1902) and his relevance to medicine in the 21st century. Curr Opin Rheumatol. 2003;15(1).
8. Firestein GS, Harris BR, McInnes I, Ruddy S, Sergent S. Vasculitis. The classification and epidemiology of systemic vasculitis. In: Saunders W, editor. Kelley´s Tetbook of Rheumatology 8th edition. 13th ed. Philadelphia; 2008.
9. McBride P. Photographs of a case of rapid destruction of the nose and face. 1897. J Laryngol Otol. 1991;105(12):1120.
10. Wegener FW. Über eine eigenartige rhinogene Granulomatose mit besonderer Beteiligung des Arterien systems und der Nieren. Beiträge zur Pathol. 1939;158:127–43.
11. DeRemee RA, McDonald TJ, Harrison EG, Coles DT. Wegener's granulomatosis. Anatomic correlates, a proposed classification. Mayo Clin Proc. 1976 Dec;51(12):777—781.
12. Godman G. Wegener's granulomatosis: pathology and review of literature. AMA Arch Pathol. 1954;58:533–53.

13. Van Der Woude FJ, Lobatto S, Permin H, Van Der Giessen M, Rasmussen N, Wiik A, et al. Autoantibodies Against Neutrophils and Monocytes: Tool for Diagnosis and Marker of Disease Activity in Wegener'S Granulomatosis. Lancet. 1985;325(8426):425–9.
14. Hutchinson J. Diseases of the arteries. Arch Surg. 1889;90(1):323.
15. Gilmour J. Giant-cell chronic arteritis. J Pathol Bacteriol. 1941;53:263–77.
16. Perings C. The American College of Rheumatology 1990 Criteria for the Classification. Dtsch Medizinische Wochenschrift. 1996;121(38):1169–75.
17. Wolfe F, Smythe HA, Yunus MB, Bennett RM, Bombardier C, Goldenberg DL, et al. The american college of rheumatology 1990 criteria for the classification of fibromyalgia. Arthritis Rheum. 1990;33(2):160–72.
18. Hunder GG, Arend WP, Bloch DA, Calabrese LH, Fauci AS, Fries JF, et al. The American College of Rheumatology 1990 Criteria for the classification of vasculitis: The introduction. Arthritis Rheum. 1990;33(8):1065–7.
19. Leavitt RY, Fauci AS, Bloch DA, Michel BA, Hunder GG, Arend WP, et al. The American College of Rheumatology 1990 criteria for the classification of wegener's granulomatosis. Arthritis Rheum. 1990;33(8):1101–7.
20. Mills JA, Michel BA, Bloch DA, Calabrese LH, Hunder GG, Arend WP, et al. The American College of Rheumatology 1990 criteria for the classification of henoch-schönlein purpura. Arthritis Rheum. 1990;33(8):1114–21.
21. Masi AT, Hunder GG, Lie JT, Michel BA, Bloch DA, Arend WP, et al. The American College of Rheumatology 1990 criteria for the classification of churg-strauss syndrome (allergic granulomatosis and angiitis). Arthritis Rheum. 1990;33(8):1094–100.
22. Jennette JC, Falk RJ, Bacon PA, Basu N, Cid MC, Ferrario F, et al. 2012 Revised International Chapel Hill consensus conference nomenclature of vasculitides. Arthritis Rheum. 2013;65(1):1–11.
23. Doubelt I, Cuthbertson D, Carette S, Khalidi NA, Koening Curry L., Langford Carol, et al. Vitamin D status in ANCA-associated vasculitis. Rheumatol Adv Pract. 2023;1–23.
24. Rao J, Allen N, Pincus T. Limitations of the 1990 American College of Rheumatology Classification Criteria in the diagnosis of vasculitis. Ann Intern Med. 1998;129(5):345–51.
25. Mohammad AJ. An update on the epidemiology of ANCA-associated vasculitis.

Rheumatol (United Kingdom). 2020;59:iii42–50.
26. Jennette JC, Falk RJ, Andrassy K, Bacon PA, Churg J, Gross WL, et al. Nomenclature of Systemic Vasculitides: Proposal of an International Consensus Conference. Arthritis Rheum. 1994;37(2):187–92.
27. Watts R, Lane S, Hanslik T, Hauser T, Hellmich B, Koldingsnes W, et al. Development and validation of a consensus methodology for the classification of the ANCA-associated vasculitides and polyarteritis nodosa for epidemiological studies. Ann Rheum Dis. 2007;66(2):222–7.
28. Johansson L, Berglin E, Eriksson O, Mohammad AJ, Dahlqvist J, Rantapää-Dahlqvist S. Complement activation prior to symptom onset in myeloperoxidase ANCA-associated vasculitis but not proteinase 3 ANCA associated vasculitis - A Swedish biobank study. Scand J Rheumatol. 2022;00(00):1–6.
29. Nilsen AT, Karlsen C, Bakland G, Watts R, Luqmani R, Koldingsnes W. Increasing incidence and prevalence of ANCA-associated vasculitis in Northern Norway. Rheumatol (United Kingdom). 2020;59(9):2316–24.
30. Craven A, Robson J, Ponte C, Grayson PC, Suppiah R, Judge A, et al. ACR/EULAR-endorsed study to develop Diagnostic and Classification Criteria for Vasculitis (DCVAS). Clin Exp Nephrol. 2013;17(5):619–21.
31. Robson JC, Grayson PC, Ponte C, Suppiah R, Craven A, Judge A, et al. 2022 American College of Rheumatology/European Alliance of Associations for Rheumatology Classification Criteria for Granulomatosis With Polyangiitis. Arthritis Rheumatol. 2022;74(3):393–9.
32. Suppiah R, Robson JC, Grayson PC, Ponte C, Craven A, Khalid S, et al. 2022 American College of Rheumatology/European Alliance of Associations for Rheumatology Classification Criteria for Microscopic Polyangiitis. Arthritis Rheumatol. 2022;74(3):400–6.
33. Grayson PC, Ponte C, Suppiah R, Robson JC, Craven A, Judge A, et al. 2022 American College of Rheumatology/European Alliance of Associations for Rheumatology Classification Criteria for Eosinophilic Granulomatosis With Polyangiitis. Arthritis Rheumatol. 2022;74(3):386–92.
34. McGlinn K, Rutherford MA, Gisslander K, Hederman L, Little MA, O'Sullivan D. FAIRVASC: A semantic web approach to rare disease registry integration. Comput Biol Med. 2022;145(February):105313.

35. Yaman B, Mcglinn K, Little MA. Towards A Rare Disease Registry Standard : Semantic Mapping of Common Data Elements Between FAIRVASC and the European Joint Programme for Rare Disease. :1–11.
36. Pop B, Fetica B, Blaga ML, Trifa AP, Achimas-Cadariu P, Vlad CI, et al. The role of medical registries, potential applications and limitations. Med Pharm Reports. 2019;92(1):7–14.
37. Nelson EC, Dixon-Woods M, Batalden PB, Homa K, Van Citters AD, Morgan TS, et al. Patient focused registries can improve health, care, and science. BMJ. 2016;354(July):1–6.
38. Norlin L, Fransson M, Eaker S, Elinder G, Litton JE. Adapting research to the 21st century - The Swedish Biobank Register. Nor Epidemiol. 2012;21(2):149–53.
39. Harron K, Dibben C, Boyd J, Hjern A, Azimaee M, Barreto ML, et al. Challenges in administrative data linkage for research. Big Data Soc. 2017;4(2).
40. Chen Z, Chen J, Collins R, Guo Y, Peto R, Wu F, et al. China Kadoorie Biobank of 0.5 million people: Survey methods, baseline characteristics and long-term follow-up. Int J Epidemiol. 2011;40(6):1652–66.
41. Lyons PA. Genome-wide association study of eosinophilic granulomatosis with polyangiitis reveals genomic loci stratified by ANCA status. Nat Commun. 2019;1–13.
42. Lyons PA, Rayner TF, Trivedi S, Holle JU, Watts RA, Jayne DRW, et al. Genetically Distinct Subsets within ANCA-Associated Vasculitis. 2012;
43. Sreih AG, Cronin K, Shaw DG, Young K, Burroughs C, Kullman J, et al. Diagnostic delays in vasculitis and factors associated with time to diagnosis. Orphanet J Rare Dis. 2021;16(1):1–8.
44. Li KJ, Semenov D, Turk M, Pope J. A meta-analysis of the epidemiology of giant cell arteritis across time and space. Arthritis Res Ther. 2021;23(1):1–10.
45. Rife E, Gedalia A. Kawasaki Disease: an Update. Curr Rheumatol Rep. 2020;22(10).
46. Burns JC, Kushner HI, Bastian JF, Shike H, Shimizu C, Matsubara T, et al. Kawasaki disease: A brief history. Pediatrics. 2000;106(2):1–8.
47. Sugiyama K, Ijiri S, Tagawa S, Shimizu K. Takayasu disease on the centenary of its discovery. Jpn J Ophthalmol. 2009;53(2):81–91.

48. Lin MT, Wu MH. The global epidemiology of Kawasaki disease: Review and future perspectives. Glob Cardiol Sci Pract. 2017;2017(3).
49. Saritas F, Donmez S, Direskeneli H, Pamuk ON. The epidemiology of Takayasu arteritis: a hospital-based study from northwestern part of Turkey. Rheumatol Int. 2016;36(7):911–6.
50. Rutter M, Bowley J, Lanyon PC, Grainge MJ. A systematic review and meta-analysis of the incidence rate of Takayasu arteritis. Rheumatol. 2021;60(11):4982–90.
51. Watts RA, Hatemi G, Burns JC, Mohammad AJ. Global epidemiology of vasculitis. Nat Rev Rheumatol. 2022;18(1):22–34.
52. Watts RA, Elane S, Scott DGI. Epidemiology of vasculitis in Europe [4]. Ann Rheum Dis. 2001;60(12):1156–7.
53. Piram M. Epidemiology of Kawasaki Disease in Europe. Front Pediatr. 2021;9(May):1–6.
54. Berti A, Cornec D, Crowson CS, Specks U, Matteson EL. The Epidemiology of Antineutrophil Cytoplasmic Autoantibody–Associated Vasculitis in Olmsted County, Minnesota: A Twenty-Year US Population–Based Study. Arthritis Rheumatol. 2017;69(12):2338–50.
55. Fujimoto S, Watts RA, Kobayashi S, Suzuki K, Jayne DRW, Scott DGI, et al. Comparison of the epidemiology of anti-neutrophil cytoplasmic antibody-associated vasculitis between Japan and the UK. Rheumatology. 2011;50(10):1916–20.
56. Nesher G, Ben-Chetrit E, Mazal B, Breuer GS. The incidence of primary systemic vasculitis in Jerusalem: A 20-year hospital-based retrospective study. J Rheumatol. 2016;43(6):1072–7.
57. Hellmich B, Lamprecht P, Spearpoint P, Götte D, Deichmann A, Buchholz I, et al. New insights into the epidemiology of ANCA-associated vasculitides in Germany: Results from a claims data study. Rheumatol (United Kingdom). 2021;60(10):4868–73.
58. Romero-Gómez C, Aguilar-García JA, García-de-Lucas MD, Cotos-Canca R, Olalla-Sierra J, García-Alegría JJ, et al. Epidemiological study of primary systemic vasculitides among adults in southern Spain and review of the main epidemiological studies. Clin Exp Rheumatol. 2015;33(2).
59. Firth J. 21 . 8 . 7 Antiglomerular basement membrane disease □. 2020;

60. Arora A, Wetter DA, Gonzalez-Santiago TM, Davis MDP, Lohse CM. Incidence of Leukocytoclastic Vasculitis, 1996 to 2010: A Population-Based Study in Olmsted County, Minnesota. Mayo Clin Proc. 2014;89(11):1515–24.
61. Srijana D, Sarosh V. Cryoglobulinemic Vasculitis. In: Cryoglobulinemic Vasculitis: Continuing Education Activity. StatPearls Publishing; 2022. p. 1.
62. Penny K, Fleming M, Kazmierczak D, Thomas A. An epidemiological study of Henoch-Schonlein purpura. Paediatr Nurs. 2010;22(10):30–5.
63. Gardner-Medwin JMM, Dolezalova P, Cummins C, Southwood TR. Incidence of Henoch-Schonlein purpura, Kawasaki disease, and rare vasculitides in children of different ethnic origins. Lancet. 2002 Oct 19;360(9341):1197–202.
64. Sjöwall C, Mandl T, Skattum L, Olsson M, Mohammad AJ. Epidemiology of hypocomplementaemic urticarial vasculitis (anti-C1q vasculitis). Rheumatol (United Kingdom). 2018;57(8):1400–7.
65. Mohammad A, Mandl T, Sturfelt G, Segelmark M. Incidence, prevalence and clinical characteristics of Behçet's disease in southern Sweden. Rheumatol (United Kingdom). 2013;52(2):304–10.
66. Lee YB, Lee SY, Choi JY, Lee JH, Chae HS, Kim JW, et al. Incidence, prevalence, and mortality of Adamantiades-Behçet's disease in Korea: a nationwide, population-based study (2006–2015). J Eur Acad Dermatology Venereol. 2018;32(6):999–1003.
67. Pugh D, Karabayas M, Basu N, Cid MC, Goel R, Goodyear CS, et al. Large-vessel vasculitis. Nat Rev Dis Prim. 2021;7(1).
68. Kitching AR, Anders HJ, Basu N, Brouwer E, Gordon J, Jayne DR, et al. ANCA-associated vasculitis. Nat Rev Dis Prim. 2020;6(1).
69. Nelveg-Kristensen KE, Szpirt W, Carlson N, McClure M, Jayne D, Dieperink H, et al. Increasing incidence and improved survival in ANCA-associated vasculitis - A Danish nationwide study. Nephrol Dial Transplant. 2022;37(1):63–71.
70. Knight A, Ekbom A, Brandt L, Askling J. Increasing incidence of Wegener's granulomatosis in Sweden, 1975-2001. J Rheumatol. 2006;33(10):2060–3.
71. Redondo-rodriguez R, Mena-vázquez N, Cabezas-lucena AM, Manrique-arija S, Mucientes A, Fernández-nebro A. Systematic Review and Metaanalysis of Worldwide Incidence and Prevalence of Antineutrophil Cytoplasmic Antibody (ANCA) Associated Vasculitis. J Clin Med. 2022;11(9).

72. Berglin E, Mohammad AJ, Dahlqvist J, Johansson L, Eriksson C, Sjöwall J, et al. Anti-neutrophil cytoplasmic antibodies predate symptom onset of ANCA-associated vasculitis. A case-control study. J Autoimmun. 2021;117.
73. Salvarani C, Cantini F, Boiardi L, Hunder GG. Review Laboratory investigations useful in giant cell arteritis and Takayasu ' s arteritis. 1967;1967(3):23–8.
74. Thompson AMS, Zanobetti A, Silverman F, Schwartz J, Coull B, Urch B, et al. Baseline repeated measures from controlled human exposure studies: Associations between ambient air pollution exposure and the systemic inflammatory biomarkers IL-6 and fibrinogen. Environ Health Perspect. 2010;118(1):120–4.
75. Hajat A, Allison M, Diez-Roux A V., Jenny NS, Jorgensen NW, Szpiro AA, et al. Long-term exposure to air pollution and markers of inflammation, coagulation, and endothelial activation a repeat-measures analysis in the multi-ethnic study of atherosclerosis (MESA). Epidemiology. 2015;26(3):310–20.
76. Yashiro M, Muso E, Itoh-Ihara T, Oyama A, Hashimoto K, Kawamura T, et al. Significantly high regional morbidity of MPO-ANCA-related angitis and/or nephritis with respiratory tract involvement after the 1995 great earthquake in Kobe (Japan). Am J Kidney Dis. 2000;35(5):889–95.
77. Farquhar HJ, McGettigan B, Chapman PT, O'Donnell JL, Frampton C, Stamp LK. Incidence of anti-neutrophil cytoplasmic antibody-associated vasculitis before and after the February 2011 Christchurch Earthquake. Intern Med J. 2017;47(1):57–61.
78. Takeuchi Y, Saito A, Ojima Y, Kagaya S, Fukami H, Sato H, et al. The influence of the Great East Japan earthquake on microscopic polyangiitis: A retrospective observational study. PLoS One. 2017;12(5):1–14.
79. Rider C V., Simmons JE. Chemical mixtures and combined chemical and nonchemical stressors: Exposure, toxicity, analysis, and risk. Chemical Mixtures and Combined Chemical and Nonchemical Stressors: Exposure, Toxicity, Analysis, and Risk. 2018. 1–556 p.
80. Patel CJ, Bhattacharya J, Butte AJ. An Environment-Wide Association Study ( EWAS ) on Type 2 Diabetes Mellitus. 2010;5(5).
81. Zheng Y, Chen Z, Pearson T, Zhao J, Hu H, Prosperi M. Design and methodology challenges of environment-wide association studies : A

systematic review. Environ Res. 2020;183(September 2019):109275.
82. Hatch M, Thomas D. Measurement issues in environmental epidemiology. Environ Health Perspect. 1993;101(SUPPL. 4):49–57.
83. Dominici F, Wang C, Crainiceanu C, Parmigiani G. Model Selection and Health Effect Estimation in Environmental Epidemiology. Epidemiology. 2008;19(4).
84. Zhong Y, Jiang C, Keung K, Zhang W, Jin Y, Hing T, et al. Environment-wide association study to identify factors associated with hematocrit : evidence from the Guangzhou Biobank Cohort Study. Ann Epidemiol. 2016;26(9):638-642.e2.
85. Hall MA, Dudek SM, Goodloe R, Crawford DC, Peissig P. Environment-wide association study (ewas) for type 2 diabetes in the marshfield personalized medicine research project biobank. Pac Symp Biocomput. 2014;200–11.
86. Lind PM, Risérus U, Salihovic S, Bavel B Van, Lind L. An environmental wide association study ( EWAS ) approach to the metabolic syndrome ☆. Environ Int. 2013;55:1–8.
87. Villiger PM, Guillevin L. Microscopic polyangiitis: Clinical presentation. Autoimmun Rev. 2010;9(12):812–9.
88. Greco A, Marinelli C, Fusconi M, Macri GF, Gallo A, De Virgilio A, et al. Clinic manifestations in granulomatosis with polyangiitis. Int J Immunopathol Pharmacol. 2016;29(2):151–9.
89. Doreille A, Buob D, Bay P, Julien M, Riviere F, Rafat C. Renal Involvement in Eosinophilic Granulomatosis With Polyangiitis. Kidney Int Reports. 2021;6(10):2718–21.
90. Palmucci S, Inì C, Cosentino S, Fanzone L, Di Pietro S, Di Mari A, et al. Pulmonary vasculitides: A radiological review emphasizing parenchymal hrct features. Diagnostics. 2021;11(12):1–16.
91. Monach PA. Biomarkers in Vasculitis. Early Hum Dev. 2007;83(4):255–62.
92. Cornec D, Gall EC Le, Fervenza FC, Specks U. ANCA-associated vasculitis-clinical utility of using ANCA specificity to classify patients. Nat Rev Rheumatol. 2016;12(10):570–9.
93. Luqmani RA. Disease assessment in systemic vasculitis. Nephrol Dial Transplant. 2015;30(February):i76–82.
94. Kobayashi S, Fujimoto S. Epidemiology of vasculitides: Differences between Japan, Europe and North America. Clin Exp Nephrol. 2013;17(5):611–4.

95. Pearce FA, Lanyon PC, Grainge MJ, Shaunak R, Mahr A, Hubbard RB, et al. Incidence of ANCA-associated vasculitis in a UK mixed ethnicity population. Rheumatol (United Kingdom). 2016;55(9):1656–63.
96. Watts RA, Gonzalez-Gay MA, Lane SE, Garcia-Porrua C, Bentham G, Scott DGI. Geoepidemiology of systemic vasculitis: Comparison of the incidence in two regions of Europe. Ann Rheum Dis. 2001;60(2):170–2.
97. Naz N, Chiu Y, Edwards M, Redfern A. Investigating whether socio-economic factors influence incidence of vasculitis in west cheshire, UK. 2019;67.
98. Aiyegbusi O, Frleta-Gilchrist M, Traynor JP, MacKinnon B, Bell S, Hunter RW, et al. ANCA-associated renal vasculitis is associated with rurality but not seasonality or deprivation in a complete national cohort study. RMD Open. 2021;7(2):1–8.
99. Alberici F, Lyons PA, Gioffredi A, Vaglio A, Martorana D, Bonatti F. Genetics of ANCA-associated vasculitides: HLA and beyond. Clin Exp Rheumatol. 2014;32(SUPPL.82):90–7.
100. Lyons PA, Rayner TF, Trivedi S, Holle JU, Watts RA, Jayne DRW, et al. Genetically Distinct Subsets within ANCA-Associated Vasculitis. N Engl J Med. 2012;367(3):214–23.
101. McDermott G, Fu X, Stone JH, Wallwork R, Zhang Y, Choi HK, et al. Association of Cigarette Smoking with Antineutrophil Cytoplasmic Antibody-Associated Vasculitis. JAMA Intern Med. 2020;02114:1–7.
102. Gómez-Puerta JA, Gedmintas L, Costenbader KH. The Association between Silica Exposure and Development of ANCA-Associated Vasculitis: Systematic Review and Meta- analysis. Autoimmun Rev. 2013;12(12):1129–1135.
103. Sarica SH, Dhaun N, Sznajd J, Harvie J, McLaren J, McGeoch L, et al. Characterizing infection in anti-neutrophil cytoplasmic antibody-associated vasculitis: Results from a longitudinal, matched-cohort data linkage study. Rheumatol (United Kingdom). 2020;59(10):3014–22.
104. Rathmann J, Jayne D, Segelmark M, Jönsson G, Mohammad AJ. Incidence and predictors of severe infections in ANCA-associated vasculitis: A population-based cohort study. Rheumatol (United Kingdom). 2021;60(6):2745–54.
105. Prendecki M, Pusey CD. Recent advances in understanding of the pathogenesis of ANCA-associated vasculitis [version 1; referees: 2 approved].

Vol. 7, F1000Research. F1000 Research Ltd; 2018.
106. Jayne DRW, Merkel PA, Schall TJ, Bekker P. Avacopan for the Treatment of ANCA-Associated Vasculitis. N Engl J Med. 2021;384(7):599–609.
107. Hoffman GS. Giant cell arteritis. Ann Intern Med. 2016;165(9):ITC65–79.
108. Smith JH, Swanson JW. Giant cell arteritis. Headache. 2014;54(8):1273–89.
109. Ness T, Bley TA, Schmidt WA, Lamprecht P. The diagnosis and treatment of giant cell arteritis. Dtsch Arztebl Int. 2013;110(21):376–84.
110. Brack A, Stanson A, Goronzy JJ, Weyand CM. LARGE-VESSEL GIANT CELL ARTERITIS. 1999;42(2):311–7.
111. Tatò F, Hoffmann U. Clinical presentation and vascular imaging in giant cell arteritis of the femoropopliteal and tibioperoneal arteries. Analysis of four cases. J Vasc Surg. 2006;44(1):176–82.
112. Martínez-Valle F, Solans-Laqué R, Bosch-Gil J, Vilardell-Tarrés M. Aortic involvement in giant cell arteritis. Autoimmun Rev. 2010;9(7):521–4.
113. Wolfe F, Smythe HA, Yunus MB, Bennett RM, Bombardier C, Goldenberg DL, et al. The american college of rheumatology 1990 criteria for the classification of giant cell arteritis. Arthritis Rheum. 1990;33(2):160–72.
114. Prieto-Peña D, Castañeda S, Martínez-Rodríguez I, Atienza-Mateo B, Blanco R, González-Gay MA. Imaging tests in the early diagnosis of giant cell arteritis. J Clin Med. 2021;10(16):1–11.
115. Koenigkam-Santos M, Sharma P, Kalb B, Oshinski JN, Weyand CM, Goronzy JJ, et al. Magnetic resonance angiography in extracranial giant cell arteritis. J Clin Rheumatol. 2011;17(6):306–10.
116. Schmidt WA. Ultrasound in the diagnosis and management of giant cell arteritis. Rheumatol (United Kingdom). 2018;57(June 2017):ii22–31.
117. de Boysson H, Aide N, Liozon E, Lambert M, Parienti JJ, Monteil J, et al. Repetitive 18F-FDG-PET/CT in patients with large-vessel giant-cell arteritis and controlled disease. Eur J Intern Med. 2017;46:66–70.
118. De Smit E, Palmer AJ, Hewitt AW. Projected worldwide disease burden from giant cell arteritis by 2050. J Rheumatol. 2015;42(1):119–25.
119. Mohammad AJ, Nilsson JÅ, Jacobsson LTH, Merkel PA, Turesson C. Incidence and mortality rates of biopsy-proven giant cell arteritis in southern Sweden. Ann Rheum Dis. 2015;74(6):993–7.
120. Garvey TD, Koster MJ, Crowson CS, Warrington KJ. Incidence, survival, and

diagnostic trends in GCA across seven decades in a North American population-based cohort. Semin Arthritis Rheum. 2021;51(6):1193–9.
121. Smeeth L, Cook C, Hall AJ. Incidence of diagnosed polymyalgia rheumatica and temporal arteritis in the United Kingdom, 1990– 2001. Ann Rheum Dis. 2006;65:1093–8.
122. Petri H, Nevitt A, Sarsour K, Napalkov P, Collinson N. Incidence of Giant Cell Arteritis and Characteristics of Patients: Data-Driven Analysis of Comorbidities. Arthritis Care Res. 2015;67(3):390–5.
123. Mollan S, Begaj I, Mackie S, Sullivan EP, Denniston A. Increase in admissions related to giant cell arteritis and polymyalgia rheumatica in the UK, 2002 - 13, without a decrease in associated sight loss: potential implications for service provision. Rheumatology. 2015;54:375–7.
124. Watts RA, Al-Taiar A, Scott DGI, MacGregor A. Prevalence and incidence of Wegener's Granulomatosis in the UK General Practice Database. Arthritis Rheum. 2009;61(10):1412–6.
125. Yates M, Graham K, Watts RA, MacGregor AJ. The prevalence of giant cell arteritis and polymyalgia rheumatica in a UK primary care population. BMC Musculoskelet Disord. 2016;17(1):1–9.
126. Greigert H, Genet C, Ramon A, Bonnotte B, Samson M. New Insights into the Pathogenesis of Giant Cell Arteritis: Mechanisms Involved in Maintaining Vascular Inflammation. J Clin Med. 2022;11(10).
127. Carmona FD, Mackie SL, Martín JE, Taylor JC, Vaglio A, Eyre S, et al. A large-scale genetic analysis reveals a strong contribution of the HLA class II region to giant cell arteritis susceptibility. Am J Hum Genet. 2015;96(4):565–80.
128. Mohan S V., Liao YJ, Kim JW, Goronzy JJ, Weyand CM. Giant cell arteritis: Immune and vascular aging as disease risk factors. Arthritis Res Ther. 2011;13(4).
129. Gloor AD, Berry GJ, Goronzy JJ, Weyand CM. Age as a risk factor in vasculitis. Semin Immunopathol. 2022;44(3):281–301.
130. Draibe J, Rodó X, Fulladosa X, Martínez-Valenzuela L, Diaz-Encarnación M, Santos L, et al. Seasonal variations in the onset of positive and negative renal ANCA-associated vasculitis in Spain. Clin Kidney J. 2018;11(4):468–73.
131. Nordborg E, Nordborg C. Giant cell arteritis: Epidemiological clues to its pathogenesis and an update on its treatment. Rheumatology. 2003;42(3):413–

21.
132. Weyand CM, Berry GJ, Goronzy JJ. The immunoinhibitory pd-1/pd-l1 pathway in inflammatory blood vessel disease. J Leukoc Biol. 2019;103(3):565–75.
133. Rodríguez-Pla A, Bosch-Gil JA, Rosselló-Urgell J, Huguet-Redecilla P, Stone JH, Vilardell-Tarres M. Metalloproteinase-2 and -9 in giant cell arteritis: Involvement in vascular remodeling. Circulation. 2005;112(2):264–9.
134. Ralston DR, Marsh CB, Lowe MP, Wewers MD. Antineutrophil cytoplasmic antibodies induce monocyte IL-8 release: Role of surface proteinase-3, α1-antitrypsin, and Fcγ receptors. J Clin Invest. 1997;100(6):1416–24.
135. Xie G, Roshandel D, Sherva R, Monach PA, Lu EY, Kung T, et al. Association of granulomatosis with polyangiitis (Wegener's) with HLA-DPB1*04 and SEMA6A gene variants: Evidence grom genome-wide analysis. Arthritis Rheum. 2013;65(9):2457–68.
136. Merkel PA, Xie G, Monach PA, Ji X, Ciavatta DJ, Byun J, et al. Identification of Functional and Expression Polymorphisms Associated With Risk for Antineutrophil Cytoplasmic Autoantibody–Associated Vasculitis. Arthritis Rheumatol. 2017;69(5):1054–66.
137. Siminovitch KA. PTPN22 and autoimmune disease. Nat Genet. 2004;36(12):1248–9.
138. Gonzalez-Gay MA. Genetic epidemiology of giant cell arteritis and polymyalgia rheumatica. Arthritis Res. 2001;3(3):154–7.
139. Weyand CM, Hicok KC, Hunder GG, Goronzy JJ. The HLA-DRB1 locus as a genetic component in giant cell arteritis. Mapping of a disease-linked sequence motif to the antigen binding site of the HLA-DR molecule. J Clin Invest. 1992;90(6):2355–61.
140. Nationa Cancer Institute. Environmental exposure [Internet]. National Institute of Health (NIH). 2023 [cited 2023 Apr 10]. p. 1.
141. US EPA. Exposure Assessment Tools by Approaches. United States Environmental Protection Agency. 2023. p. 1.
142. US EPA. Exposure Assessment Tools by Approaches - Direct Measurement (Point-of-Contact Measurement) [Internet]. United States Environmental Protection Agency. 2023 [cited 2023 Apr 11]. p. 1.
143. Dybwad M, Skogan G, Blatny JM. Comparative testing and evaluation of nine different air samplers: End-to-end sampling efficiencies as specific

performance measurements for bioaerosol applications. Aerosol Sci Technol. 2014;48(3):282–95.
144. US EPA. Exposure Assessment Tools by Approaches - Indirect Estimation (Scenario Evaluation). United state Environmental Protection Agency. 2023. p. 1.
145. Weiner M, Bjørneklett R, Hru Z. Proteinase-3 and myeloperoxidase serotype in relation to demographic factors and geographic distribution in anti-neutrophil cytoplasmic antibody-associated glomerulonephritis. Nephrol Dial Transplant. 2018;(April 2018):301–8.
146. Gatenby PA, Lucas RM, Engelsen O, Ponsonby AL, Clements M. Antineutrophil cytoplasmic antibody-associated vasculitides: Could geographic patterns be explained by ambient ultraviolet radiation? Arthritis Care Res. 2009;61(10):1417–24.
147. Scott J, Havyarimana E, Navarro-Gallinad A, White A, Wyse J, van Geffen J, et al. The association between ambient UVB dose and ANCA-associated vasculitis relapse and onset. Arthritis Res Ther. 2022;24(1):1–14.
148. Watts RA, Scott DG. ANCA vasculitis over the world. What do we learn from country differences? Press Medicale. 2013;42(4 PART2):591–3.
149. Pearce FA, Craven A, Merkel PA, Luqmani RA, Watts RA. Global ethnic and geographic differences in the clinical presentations of anti-neutrophil cytoplasm antibody-associated vasculitis. Rheumatol (United Kingdom). 2017;56(11):1962–9.
150. Li J, Cui Z, Long JY, Huang W, Wang JW, Wang H, et al. The frequency of ANCA-associated vasculitis in a national database of hospitalized patients in China. Arthritis Res Ther. 2018;20(1):226.
151. Wing S, Rider LG, Johnson JR, Miller FW, Matteson EL, Crowson CS, et al. Do solar cycles influence giant cell arteritis and rheumatoid arthritis incidence? BMJ Open. 2015;5(5):1–9.
152. Rodó X, Ballester J, Cayan D, Melish ME, Nakamura Y, Uehara R, et al. Association of Kawasaki disease with tropospheric wind patterns. Sci Rep. 2011;1(April 1979):1–7.
153. Stamatis P, Turesson C, Nilsson JÅ, Mohammad A. 164. Incidence and Seasonal Variation of Biopsy-Proven Giant Cell Arteritis – Revisited a 20 Years Population-Based Study From Sweden. Rheumatology.

2019;58(Supplement_2).
154. Gokoffski KK, Chatterjee A, Khaderi SK. Seasonal incidence of biopsy-proven giant cell arteritis: a 20-year retrospective study of the University of California Davis Medical System. Clin Exp Rheumatol. 2019;37 Suppl 1(2):90–7.
155. Kønig EB, Stormly Hansen M, Foldager J, Siersma V, Loft A, Terslev L, et al. Seasonal variation in biopsy-proven giant cell arteritis in Eastern Denmark from 1990-2018. Acta Ophthalmol. 2021;99(5):527–32.
156. Raheel S, Crowson CS, Matteson EL, Joseph S. Seasonal Variation in Incidence of Giant Cell Arteritis : A Population-Based Cohort Study. In: ACR/ARHP Annual Meeting. 2018. p. 1–2.
157. Frleta-gilchrist M, Aiyegbusi O, Mackinnon M, Taynor J, McQuarrie E, Stevens K, et al. Seasonality has no association with timing of diagnosis of renal ANCA associated vasculitis. Nephrol Dial Transplant. 2018;1093.
158. Chung EY, Risi D, Holt JL, Lonergan M, Kotwal S, Yong K, et al. A retrospective study on the epidemiology of ANCA -associated vasculitis in two Australian health districts . Intern Med J. 2020;
159. Aiyegbusi O, Frleta-Gilchrist M, Traynor JP, MacKinnon B, Bell S, Hunter RW, et al. ANCA-associated renal vasculitis is associated with rurality but not seasonality or deprivation in a complete national cohort study. RMD Open. 2021;7(2):e001555.
160. Koldingsnes W, Nossent H. Epidemiology of Wegener's granulomatosis in northern Norway. Arthritis Rheum. 2000;43(11):2481–7.
161. Tidman M, Olander R, Svalander C, Danielsson D. Patients hospitalized because of small vessel vasculitides with renal involvement in the period 1975-95: Organ involvement, anti-neutrophil cytoplasmic antibodies patterns, seasonal attack rates and fluctuation of annual frequencies. J Intern Med. 1998;244(2):133–41.
162. Draibe J, Rodo X, Fulladosa X, Martinez-Valenzuela L, Diaz-Encarnacion M, Santos L, et al. Seasonal variations in the onset of positive and negative renal ANCA-associated vasculitis in Spain. Clin Kidney J. 2018;11(4):468–73.
163. Caballero-Islas AE, Hoyo-Ulloa I, García-Castro A, Hinojosa-Azaola A. Severe infections in patients with anti-neutrophil cytoplasmic antibody-associated vasculitis: a retrospective cohort study with a clinical phenotype approach. Rheumatol Int. 2020;40(10):1657–66.

164. Kakoullis L, Parperis K, Papachristodoulou E, Panos G. Infection-induced myeloperoxidase specific antineutrophil cytoplasmic antibody (MPO-ANCA) associated vasculitis: A systematic review. Clin Immunol. 2020;220(August):108595.
165. Mitchison RS, Rees AJ, Pussell BA, Lockwood CM, Peters DK. Relapses in Wegener's granulomatosis: The role of infection. Br Med J. 1980;281(6244):836–8.
166. Nielsen PR, Kragstrup TW, Deleuran BW, Benros ME. Infections as risk factor for autoimmune diseases – A nationwide study. J Autoimmun. 2016;74:176–81.
167. Bonetto C, Trotta F, Felicetti P, Alarcón GS, Santuccio C, Bachtiar NS, et al. Vasculitis as an adverse event following immunization – Systematic literature review. Vaccine. 2016;34(51):6641–51.
168. Jeffs LS, Nitschke J, Tervaert JWC, Peh CA, Hurtado PR. Viral RNA in the influenza vaccine may have contributed to the development of ANCA-associated vasculitis in a patient following immunisation. Clin Rheumatol. 2016;35(4):943–51.
169. Felicetti P, Trotta F, Bonetto C, Santuccio C, Brauchli Pernus Y, Burgner D, et al. Spontaneous reports of vasculitis as an adverse event following immunization: A descriptive analysis across three international databases. Vaccine. 2016;34(51):6634–40.
170. Mahr A, Guillevin L, Poissonnet M, Aymé S. Prevalences of Polyarteritis Nodosa, Microscopic Polyangiitis, Wegener's Granulomatosis, and Churg-Strauss Syndrome in a French Urban Multiethnic Population in 2000: A Capture-Recapture Estimate. Arthritis Care Res. 2004;51(1):92–9.
171. Anderson K, Klassen J, Stewart SA, Taylor-Gjevre RM. Does geographic location affect incidence of ANCA-associated renal vasculitis in Northern Saskatchewan, Canada? Rheumatol (United Kingdom). 2013;52(10):1840–4.
172. Ormerod AS, Cook MC. Epidemiology of primary systemic vasculitis in the Australian Capital Territory and south-eastern New South Wales. Intern Med J. 2008;38(11):816–23.
173. Haubitz M, Woywodt A, De Groot K, Haller H, Goebel U. Smoking habits in patients diagnosed with ANCA associated small vessel vasculitis. Ann Rheum Dis. 2005;64(10):1500–2.
174. World Meteorological Organization. Airborne Dust: A Hazard to Human Health,

Environment and Society. WMO. 2020. p. 1.
175. Webber MP, Moir W, Zeig-Owens R, Glaser MS, Jaber N, Hall C, et al. Nested case-control study of selected systemic autoimmune diseases in world trade center rescue/recovery workers. Arthritis Rheumatol. 2015;67(5):1369–76.
176. Ying D, Schmajuk G, Trupin L, Blanc PD. Inorganic Dust Exposure During Military Service as a Predictor of Rheumatoid Arthritis and Other Autoimmune Conditions. ACR Open Rheumatol. 2021;3(7):466–74.
177. Ishikawa Y, Terao C. The impact of cigarette smoking on risk of rheumatoid arthritis: A narrative review. Vol. 9, Cells. 2020.
178. Parisis D, Bernier C, Chasset F, Arnaud L. Impact of tobacco smoking upon disease risk, activity and therapeutic response in systemic lupus erythematosus: A systematic review and meta-analysis. Autoimmun Rev. 2019;18(11):102393.
179. Brennan DN, Ungprasert P, Warrington KJ, Koster MJ. Smoking as a risk factor for giant cell arteritis: A systematic review and meta-analysis. Semin Arthritis Rheum. 2018;48(3):529–37.
180. Benarous L, Terrier B, Puéchal X, Dunogué B, Cohen P, Le Jeunne C, et al. Tobacco differentially affects the clinical-biological phenotypes of ANCA-associated vasculitides. Clin Exp Rheumatol. 2015;33(2).
181. Requia WJ, Adams MD, Arain A, Papatheodorou S, Koutrakis P, Mahmoud M. Global Association of Air Pollution and Cardiorespiratory Diseases: A Systematic Review, Meta-Analysis, and Investigation of Modifier Variables. Am J Public Health. 2018;108:S123–30.
182. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev. 2015;4(1):1.
183. Gregorini G, Ferioli A, Tardanico R, Lancini L, Maiorca R. Association between silica exposure and necrotizing crescentic glomerulonephritis with p-anca and anti-mpo antibodies: a hospital-based case-control study. Adv Exp Med Biol. 1993;336:435–40.
184. Nuyts GD, Van Vlem E, De Vos A, Daelemans RA, Rorive G, Elseviers MM, et al. Wegener granulomatosis is associated to exposure to silicon compounds: a case-control study. Nephrol Dial Transplant. 1995;10(7):1162–5.
185. Sutton-Tyrrell K. Assessing bias in case-control studies. Stroke.

1991;22(7):938–42.
186. Maritati F, Peyronel F, Fenaroli P, Pegoraro F, Lastrucci V, Benigno GD, et al. Occupational Exposures and Smoking in Eosinophilic Granulomatosis With Polyangiitis: A Case–Control Study. Arthritis Rheumatol. 2021;73(9):1694–702.
187. Stamp LK, Chapman PT, Francis J, Beckert L, Frampton C, Watts RA, et al. Association between environmental exposures and granulomatosis with polyangiitis in Canterbury, New Zealand. Arthritis Res Ther. 2015;17(1):333.
188. Willeke P, Schlüter B, Sauerland C, Becker H, Reuter S, Jacobi A, et al. Farm exposure as a differential risk factor in ANCA-associated vasculitis. PLoS One. 2015;10(9):e0137196.
189. Rodó X, Curcoll R, Robinson M, Ballester J, Burns JC, Cayan DR, et al. Tropospheric winds from northeastern China carry the etiologic agent of Kawasaki disease from its source to Japan. Proc Natl Acad Sci U S A. 2014;111(22):7952–7.
190. Rodó X, Ballester J, Cayan D, Melish ME, Nakamura Y, Uehara R, et al. Association of Kawasaki disease with tropospheric wind patterns. Sci Rep. 2011;1.
191. Jorquera H, Borzutzky A, Hoyos-Bachiloglu R, García A. Association of Kawasaki disease with tropospheric winds in Central Chile: Is wind-borne desert dust a risk factor? Environ Int. 2015;78:32–8.
192. Abdulkader R, Lane SE, Scott DGI, Watts RA. Classification of vasculitis: EMA classification using CHCC 2012 definitions. Ann Rheum Dis. 2013;72(11):1888.
193. Worlds Health Organization (WHO). Air quality and health [Internet]. 2022. [cited 2022 Nov 20]. p. 1.
194. European Environment Agency. Air quality in Europe. Vol. 1. 2019. 83 p.
195. United State Environmental Protection Agency. Air Pollution: Current and Future Challenges. US EPA. 2022. p. 1.
196. United State Environmental Protection Agency. Criteria Air Pollutants [Internet]. US EPA. 2022. p. 1.
197. Sun G, Hazlewood G, Bernatsky S, Kaplan GG, Eksteen B, Barnabe C. Association between Air Pollution and the Development of Rheumatic Disease: A Systematic Review. Int J Rheumatol. 2016;2016:5356307.
198. Di D, Zhang L, Wu X, Leng R. Long-term exposure to outdoor air pollution and

the risk of development of rheumatoid arthritis: A systematic review and meta-analysis. Semin Arthritis Rheum. 2020;50(2):266–75.
199. Gilcrease GW, Padovan D, Heffler E, Peano C, Massaglia S, Roccatello D, et al. Is air pollution affecting the disease activity in patients with systemic lupus erythematosus? State of the art and a systematic literature review. Eur J Rheumatol. 2020;7(1):31–4.
200. International Labour Organization (ILO). International standard Occupational Classification (ISOC). ONS. 2008. p. 1.
201. Sadhra SS, Kurmi OP, Chambers H, Lam KBH, Fishwick D, Cullinan P, et al. Development of an occupational airborne chemical exposure matrix. Occup Med (Chic Ill). 2016;66(5):358–64.
202. Descatha A, Fadel M, Sembajwe G, Peters S, Evanoff BA. Job-Exposure Matrix: A Useful Tool for Incorporating Workplace Exposure Data Into Population Health Research and Practice. Front Epidemiol. 2022;2(April):1–7.
203. The Joanna Briggs Institute. Checklist for Case Control Studies. Joanna Briggs Inst Crit Apprais tools. 2016;1–6.
204. Moola S, Munn Z, Tufanaru C, Aromataris E, Sears K, Sfetcu R, Currie M, Lisy K, Qureshi R, Mattis P, Mu PMoola S, Munn Z, Tufanaru C, Aromataris E, Sears K, Sfetcu R, Currie M, Lisy K, Qureshi R, Mattis P MP. Chapter 7: Systematic reviews of etiology and risk. In: Aromataris E, Munn Z (Editors). JBI Manual for Evidence Synthesis. 2020.
205. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986;7(3):177–88.
206. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21(11):1539–58.
207. Higgins JPT, Thompson SG, Deeks J, Altman DG. Measuring inconsistency in knowledgebases. BMJ. 2003;557–60.
208. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. Br Med J. 1997;315(7109):629–34.
209. Pearce F, Lanyon P, Grainge M, Reena S, Alfred M, Richard H, et al. Incidence of ANCA-associated vasculitis in a UK mixed ethnicity population. Rheumatology (Oxford). 2016;55(9).
210. Albert D, Clarkin C, Komoroski J, Brensinger CM, Berlin JA. Wegener's granulomatosis: Possible role of environmental agents in its pathogenesis.

Arthritis Rheum. 2004;51(4):656–64.
211. Hogan SL, Cooper GS, Savitz DA, Nylander-French LA, Parks CG, Chin H, et al. Association of silica exposure with anti-neutrophil cytoplasmic autoantibody small-vessel vasculitis: a population-based, case-control study. Clin J Am Soc Nephrol. 2007;2(2):290–9.
212. Beaudreuil S, Lasfargues G, Laueriere L, El Ghoul Z, Fourquet F, Longuet C, et al. Occupational exposure in ANCA-positive patients: A case-control study. Kidney Int. 2005;67(5):1961–6.
213. Lane SE, Watts RA, Bentham G, Innes NJ, Scott DGI. Are environmental factors important in primary systemic vasculitis? A case-control study. Arthritis Rheum. 2003;48(3):814–23.
214. Breuer GS, Nesher G, Nesher R. Rate of discordant findings in bilateral temporal artery biopsy to diagnose giant cell arteritis. J Rheumatol. 2009;36(4):794–6.
215. Boyev LR, Miller NR, Green WR. Efficacy of unilateral versus bilateral temporal artery biopsies for the diagnosis of giant cell arteritis. Am J Ophthalmol. 1999;128(2):211–5.
216. Hamzah Mahmood-Rao, Peter Lanyon RW and FP. Is anca-associated vasculitis more common in areas with more polluted air? Rheumatology. 2018;(May):4971253.
217. González-Gay MA, Garcia-Porrua C, Rivas MJ, Rodriguez-Ledo P, Llorca J. Epidemiology of biopsy proven giant cell arteritis in northwestern Spain: Trend over an 18 year period. Ann Rheum Dis. 2001;60(4):367–71.
218. Richier Q, Deltombe T, Foucher A, Roussin C, Randrianjohany A, Gerber A, et al. Giant cell arteritis epidemiology in La reunion: a retrospective cases series. In: Vasculitis. BMJ Annals of the Rheumatic Diseases; 2018. p. 1488.2-1489.
219. Gonzalez-Gay MA, Miranda-Filloy JA, Lopez-Diaz MJ, Perez-Alvarez R, Gonzalez-Juanatey C, Sanchez-Andrade A, et al. Giant cell arteritis in northwestern Spain: A 25-year epidemiologic study. Medicine (Baltimore). 2007;86(2):61–8.
220. Kisza K, Murchison AP, Bilyk JR, Eagle RC, Sergott R, Dai Y, et al. Giant cell arteritis incidence: Analysis by season and year in mid-Atlantic United States. Clin Exp Ophthalmol. 2013;41(6):577–81.
221. Stamatis P, Turesson C, Nilsson JA ke, Mohammad A. Incidence And

Seasonal Variation Of Biopsy-Proven Giant Cell Arteritis – Revisited: A 20-Year Population-Based Study From Sweden. Epidemiol ii71 164. 2016;164.
222. Gokoffski KK, Chatterjee A, Khaderi SK. Seasonal incidence of biopsy-proven giant cell arteritis: a 20-year retrospective study of the University of California Davis Medical System. Clin Exp Rheumatol. 2019;37(2):90–7.
223. Lane SE, Watts RA, Scott DGI. Seasonal variations in onset of Wegener's granulomatosis: Increased in summer? [3]. J Rheumatol. 2007;34(4):889–90.
224. Petursdottir V, Johansson H, Nordborg E, Nordborg C. The epidemiology of biopsy-positive giant cell arteritis: Special reference to cyclic fluctuations. Rheumatology. 1999;38(12):1208–12.
225. Parks CG, Cooper GS, Nylander-French LA, Storm JF, Archer JD. Assessing exposure to crystalline silica from farm work: A population-based study in the southeastern United States. Ann Epidemiol. 2003;13(5):385–92.
226. Boudigaard SH, Schlünssen V, Vestergaard JM, Søndergaard K, Torén K, Peters S, et al. Occupational exposure to respirable crystalline silica and risk of autoimmune rheumatic diseases: A nationwide cohort study. Int J Epidemiol. 2021;50(4):1213–26.
227. Blanc PD, Järvholm B, Torén K. Prospective risk of rheumatologic disease associated with occupational exposure in a cohort of male construction workers. Am J Med. 2015;128(10):1094–101.
228. Steenland K, Sanderson W, Calvert GM. Kidney disease and arthritis in a cohort study of workers exposed to silica. Epidemiology. 2001;12(4):405–12.
229. Khelgi V, Gupta A, Venuthurapalli S, Lee A, Govindarajulu S. Prevalence of ANCA vasculitis in DDHHS population. Nephrology. 2017;22(Supplement 3):62.
230. van Iterson M, van Zwet EW, Heijmans BT, 't Hoen PAC, van Meurs J, Jansen R, et al. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. Genome Biol. 2017;18(1):1–13.
231. Bru T. Assessment of exposure in epidemiological studies : the example of silica dust. 2008;452–61.
232. Davis JL, Wernicke BP, Tamisiea ME. On seasonal signals in geodetic time series. J Geophys Res Solid Earth. 2012;117(1):1–10.
233. Pearce FA, McGrath C, Sandhu R, Packham J, Watts RA, Rhodes B, et al.

Outcomes and compliance with standards of care in anti-neutrophil cytoplasmic antibody-associated vasculitis-insights from a large multiregion audit. Rheumatol Adv Pract. 2018;2(2):1–7.

234. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLoS Med. 2015;12(3):1–10.
235. Elliott P, Peakman TC. The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. Int J Epidemiol. 2008;37(2):234–44.
236. Burton PR, Hansell AL, Fortier I, Manolio TA, Khoury MJ, Little J, et al. Size matters: Just how big is BIG?: Quantifying realistic sample size requirements for human genome epidemiology. Int J Epidemiol. 2009;38(1):263–73.
237. Sarkar C, Webster C, Gallacher J. UK Biobank Urban Morphometric Platform (UKBUMP) – a nationwide resource for evidence-based healthy city planning and public health interventions. Ann GIS. 2015;21(2):135–48.
238. Elias P, Mcknight A, Davies R, Birch M. Standard Occupational Classification 2000. Volume 1. Structure and description of unit groups. 2000;1.
239. Booth M. Assessment of physical activity: An international perspective. Res Q Exerc Sport. 2000;71:114–20.
240. Harron K, Dibben C, Boyd J, Hjern A, Azimaee M, Barreto ML, et al. Challenges in administrative data linkage for research. 2017;(December):1–12.
241. Farr W. Letter to the Registrar-General on the mortality in the registration districts of England during the years 1861–70 in Supplement to the 35 th annual report of the Registrar-General of births, deaths, and marriages in England [Internet]. London; 1975.
242. Acheson ED. Oxford record linkage study a central file of morbidity and mortality records for a pilot population. Vol. 18, Brit. J. prev. soc. Med. 1964.
243. Dunn H. Record Linkage. Am J Public Health. 1946;36:1412–6.
244. Marshall JT. Canada's national vital statistics index. Popul Stud (NY). 1947;1(2):204–11.
245. Newcombe HB, ~kennedy~ JM. Record Linkage Making Maximum Use of the Discriminating Power of Identifying Information.
246. Newcombe HB, Kennedy JM, Axford SJ, James AP. Automatic linkage of vital

records. Science (80- ). 1959;130(3381):954–9.
247. Shapiro J. The NHS: the story so far (1948–2010) [Internet]. 2010.
248. Chamnan P, Simmons RK, Hori H, Sharp S, Khaw KT, Wareham NJ, et al. A simple risk score using routine data for predicting cardiovascular disease in primary care. Br J Gen Pract. 2010;60(577):327–34.
249. Clair L, Anderson H, Anderson C, Ekuma O, Prior HJ. Cardiovascular disease and the risk of dementia: a survival analysis using administrative data from Manitoba. Can J Public Heal. 2022;113(3):455–64.
250. Garg AX, Prasad GVR, Thiessen-Philbrook HR, Ping L, Melo M, Gibney EM, et al. Cardiovascular disease and hypertension risk in living kidney donors: An analysis of health administrative data in Ontario, Canada. Transplantation. 2008;86(3):399–406.
251. Holman CDAJ, Bass AJ, Rosman DL, Smith MB, Semmens JB, Glasson EJ, et al. A decade of data linkage in Western Australia : strategic design , applications and benefits of the WA data linkage system. 2008;32(4):766–77.
252. Fair Statistics Canada M. Generalized Record Linkage System-Statistics Canada's Record Linkage Software. Vol. 33, Austrian Journal of Statistics. 2004.
253. Schnell R, Bachteler T, Reiher J. Privacy-preserving record linkage using Bloom filters. BMC Med Informatics Decis Mak Privacy-preserving Rec Link using Bloom filters. 2009;11:1–11.
254. Dibben C, Elliot M, Gowans H, Lightfoot D. Methodological Developments in Data Linkage. In: The Data Linkage Environment. London: John Wiley & Sons Ltd; 2015. p. 36–62.
255. Information Commissioner's Office. Data Protection Act 2018 [Internet]. London; 2018.
256. Adamska L, Allen N, Flaig R, Sudlow C, Lay M, Landray M. Challenges of linking to routine healthcare records in UK Biobank. Trials. 2015;16(S2):O68.
257. Nosrati E. Comment Harnessing administrative data to study health inequality. Lancet Public Heal. 2022;7(9):e726–7.
258. Glasgow City Alcohol and Drug Partnership. Glasgow City Alcohol and Drug Strategy [Internet]. Glasgow; 2020.
259. DEFRA. Clean air strategy [Internet]. 2019.
260. Moorthie S, Hayat S, Zhang Y, Parkin K, Philips V, Bale A, et al. Rapid

systematic review to identify key barriers to access, linkage, and use of local authority administrative data for population health research, practice, and policy in the United Kingdom. BMC Public Health. 2022;22(1):1–13.
261. Playford CJ, Gayle V, Connelly R, Gray AJG. Administrative social science data : The challenge of reproducible research. 2016;(December):1–13.
262. Information N, Ireland N. Accessing electronic administrative health data for research takes time. 2013;98(5):2–3.
263. Neter J, Maynes S, Ramathan R. The Effect of Mismatching on the Measurement of Response Errors. J Am Stat Assoc. 1965;60(312):2–8.
264. UK Government. Clean Air Act 1993. London: Parliament: House of Commons; 1993 p. c.11.
265. Cromar KR, Duncan BN, Bartonova A, Benedict K, Brauer M, Habre R, et al. Air pollution monitoring for health research and patient care: An official American Thoracic Society workshop report. Ann Am Thorac Soc. 2019;16(10):1207–14.
266. Atkinson RW, Carey IM, Kent AJ, Van Staa TP, Anderson HR, Cook DG. Long-term exposure to outdoor air pollution and the incidence of chronic obstructive pulmonary disease in a national English cohort. Occup Environ Med. 2015;72(1):42–8.
267. Morbey RA, Elliot AJ, Harcourt S, Smith S, De Lusignan S, Pebody R, et al. Estimating the burden on general practitioner services in England from increases in respiratory disease associated with seasonal respiratory pathogen activity. Epidemiol Infect. 2018;146(11):1389–96.
268. Moan J, Lagunova Z, Bruland Ø, Juzeniene A. Seasonal variations of cancer incidence and prognosis. Dermatoendocrinol. 2010;2(2):55–7.
269. Huang G, Lee D, Scott EM. Multivariate space-time modelling of multiple air pollutants and their health effects accounting for exposure uncertainty. Stat Med. 2018;37(7):1134–48.
270. UK department for Environment Food and Rural Affairs. Modelled background pollution data [Internet]. UK department for Environment Food and Rural Affairs. 2020 [cited 2022 Jul 21].
271. Lee EK, Romeiko XX, Zhang W, Feingold BJ, Khwaja HA, Zhang X, et al. Residential Proximity to Biorefinery Sources of Air Pollution and Respiratory Diseases in New York State. Environ Sci Technol. 2021;55(14):10035–45.

272. O'Connor AM, Auvermann BW, Dzikamunhenga RS, Glanville JM, Higgins JPT, Kirychuk SP, et al. Updated systematic review: Associations between proximity to animal feeding operations and health of individuals in nearby communities. Syst Rev. 2017;6(1).
273. Hart JE, Laden F, Puett RC, Costenbader KH, Karlson EW. Exposure to traffic pollution and increased risk of rheumatoid arthritis. Environ Health Perspect. 2009;117(7):1065–9.
274. Hoek G, Beelen R, de Hoogh K, Vienneau D, Gulliver J, Fischer P, et al. A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmos Environ. 2008;42(33):7561–78.
275. Cabaneros SM, Calautit JK, Hughes BR. A review of artificial neural network models for ambient air pollution prediction. Environ Model Softw. 2019;119(June 2018):285–304.
276. Brown A, Marshall DR. A basic sampling strategy: Theory and practice [Internet]. Collecting Plant Genetic Diversity: Technical Guidelines. 1995. p. 75–91.
277. Hollis D, McCarthy M, Kendon M, Legg T, Simpson I. HadUK-Grid—A new UK dataset of gridded climate observations. Geosci Data J. 2019;6(2):151–9.
278. US EPA. A Guide to Siting and Installing Air Sensors. Environmental protection agency of US. 2020. p. 1.
279. Guo R, Qi Y, Zhao B, Pei Z, Wen F, Wu S, et al. High-Resolution Urban Air Quality Mapping for Multiple Pollutants based on Dense Monitoring Data and Machine Learning. Int J Environ Res Public Health. 2022;19(13).
280. Karagulian F, Gerboles M, Barbiere M, Kotsev A, Lagler F, Borowiak A. Review of sensors for air quality monitoring [Internet]. JRC Technical Reports, European Commission. 2019. 54 p.
281. Hanigan I, Hall G, Dear KBG. A comparison of methods for calculating population exposure estimates of daily weather for health research. 2006;16:1–16.
282. Bell ML. The use of ambient air quality modeling to estimate individual and population exposure for human health research: A case study of ozone in the Northern Georgia Region of the United States. Environ Int. 2006;32(5):586–93.
283. Reddy BP, Houlding B, Hederman L, Canney M, Debruyne C, O'Brien C, et al. Data linkage in medical science using the resource description framework: the

AVERT model. HRB Open Res. 2019;1:20.
284. Zhu X, She B, Guo W, Bao S, Chen D. Integrating spatial data linkage and analysis services in a geoportal for China Urban Research. Trans GIS. 2015;19(1):107–28.
285. Kenward RE, Clarke RT, Hodder KH, Walls SS. Density and linkage estimators of home range: Nearest-neighbor clustering defines multinuclear cores. Ecology. 2001;82(7):1905–20.
286. Okabe A, Boots B, Sugihara K. Nearest neighbourhood operations with generalized voronoi diagrams: A review. Int J Geogr Inf Syst. 1994;8(1):43–71.
287. Andersson K, Bray F, Arbyn M, Storm H, Zanetti R, Hallmans G, et al. The interface of population-based cancer registries and biobanks in etiological and clinical research – current and future perspectives. 2010;
288. Huang Y, Zhu M, Ji M, Fan J, Xie J, Wei X, et al. Air pollution, genetic factors, and the risk of lung cancer. A prospective study in the UK biobank. Am J Respir Crit Care Med. 2021;204(7):817–25.
289. Warkentin MT, Lam S, Hung RJ. Determinants of impaired lung function and lung cancer prediction among never-smokers in the UK Biobank cohort. EBioMedicine. 2019;47:58–64.
290. Wang J, Lin C, Chu Y, Deng H, Shen Z. Association between long-term exposure to air pollution and the risk of incident laryngeal cancer: a longitudinal UK Biobank-based study. Environ Sci Pollut Res. 2023;58295–303.
291. Pukkala E. Biobanks and Registers in Epidemiologic Research on Cancer. Vol. 675.
292. Dillner J. Methods in Biobanking. [Internet]. Vol. 675, Methods in Molecular Biology. Springer; 2011. 307–312 p.
293. Kollhorst B, Reinders T, Grill S, Eberle A, Intemann T, Kieschke J, et al. Record linkage of claims and cancer registries data—Evaluation of a deterministic linkage approach based on indirect personal identifiers. Pharmacoepidemiol Drug Saf. 2022;31(12):1287–93.
294. DEFRA. Statistical Digest of Rural England Population. Gov Stat Serv. 2021;(October).
295. Gokuladhas S, Schierding W, Golovina E, Fadason T, O'Sullivan J. Unravelling the Shared Genetic Mechanisms Underlying 18 Autoimmune Diseases Using a

Systems Approach. Front Immunol. 2021;12(August):1–13.
296. Kuiper JJW, Prinz JC, Stratikos E, Kuśnierczyk P, Arakawa A, Springer S, et al. EULAR study group on ‘ MHC- - opathy ’: identifying overarching mechanisms across disciplines and borders. 2023;(figure 1):1–10.
297. Smolen JS, Aletaha D, Barton A, Burmester GR, Emery P, Firestein GS, et al. Rheumatoid arthritis. Nat Rev Dis Prim. 2018;4:1–23.
298. Zhang HG, McDermott G, Seyok T, Huang S, Dahal K, L’Yi S, et al. Identifying shared genetic architecture between rheumatoid arthritis and other conditions: a phenome-wide association study with genetic risk scores. eBioMedicine. 2023;92:104581.
299. Carmona FD, González-Gay MA, Martín J. Genetic component of giant cell arteritis. Rheumatol (United Kingdom). 2014;53(1):6–18.
300. Chen S, Pu W, Guo S, Jin L, He D, Wang J. Genome-wide DNA methylation profiles reveal common epigenetic patterns of interferon-related genes in multiple autoimmune diseases. Front Genet. 2019;10(APR).
301. Alexeeff SE, Coull BA, Gryparis A, Suh H, Sparrow D, Vokonas PS, et al. Medium-term exposure to traffic-related air pollution and markers of inflammation and endothelial function. Environ Health Perspect. 2011;119(4):481–6.
302. Forbes LJL, Patel MD, Rudnicka AR, Cook DG, Bush T, Stedman JR, et al. Chronic exposure to outdoor air pollution and markers of systemic inflammation. Epidemiology. 2009;20(2):245–53.
303. Pope CA, Hansen ML, Long RW, Nielsen KR, Eatough NL, Wilson WE, et al. Ambient particulate air pollution, heart rate variability, and blood markers of inflammation in a panel of elderly subjects. Environ Health Perspect. 2004;112(3):339–45.
304. Viehmann A, Hertel S, Fuks K, Eisele L, Moebus S, Möhlenkamp S, et al. Long-term residential exposure to urban air pollution, and repeated measures of systemic blood markers of inflammation and coagulation. Occup Environ Med. 2015;72(9):656–63.
305. DEFRA. Air Pollution in UK. 2003;
306. Andy Cook, Andrew Kent, Mark Harrison CW and HW. UK Air Quality Forecasting : Annual Report 2005 UK Air Quality Forecasting : Annual Report 2005. 2005;(1).

307. DEFRA. Air Pollution in the UK 2015 [Internet]. 2016.
308. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B. 1995;57(1):289–300.
309. Benno Haarman BCM, Riemersma-Van der Lek RF, Nolen WA, Mendes R, Drexhage HA, Burger H. Feature-expression heat maps - A new visual method to explore complex associations between two variable sets. J Biomed Inform. 2015;53:156–61.
310. Edwards M, Naz N, Redfern A, Chiu Y. Investigating Whether Socio-Economic Factors Influence Incidence of Vasculitis in Westchire, UK. 2016;(June):5257.
311. Watts RA, Mooney J, Scott DGI. 339. The Epidemiology of Vasculitis in the UK. Rheumatology. 2014;53(suppl_1):i187–i187.
312. Watts RA, Al-Taiar A, Scott DGI, MacGregor A. Prevalence and Incidence of Wegner's Granulomatosis in the UK General Practice. Arthritis Rheum. 1959;61(1):1412–6.
313. C.-R. J, H.-Y. H, Jung CR, Hsieh HY, Hwang BF. Air Pollution as a Potential Determinant of Rheumatoid Arthritis: A Population-based Cohort Study in Taiwan. Epidemiology. 2017;28(Supplement 1):S54–9.
314. Hart JE, Källberg H, Laden F, Bellander T, Costenbader KH, Holmqvist M, et al. Ambient air pollution exposures and risk of rheumatoid arthritis: Results from the Swedish EIRA case-control study. Ann Rheum Dis. 2013;72(6):888–94.
315. Zhao CN, Xu Z, Wu GC, Mao YM, Liu LN, Qian-Wu, et al. Emerging role of air pollution in autoimmune diseases. Autoimmun Rev. 2019;18(6):607–14.
316. Lanki T, Hampel R, Tiittanen P, Andrich S, Beelen R, Brunekreef B, et al. Air pollution from road traffic and systemic inflammation in adults: A cross-sectional analysis in the European ESCAPE project. Environ Health Perspect. 2015;123(8):785–91.
317. Bernatsky S, Fournier M, Pineau CA, Clarke AE, Vinet E, Smargiassi A. Associations between ambient fine particulate levels and disease activity in patients with systemic lupus erythematosus (SLE). Environ Health Perspect. 2011;119(1):45–9.
318. Zhao N, Smargiassi A, Hatzopoulou M, Colmegna I, Hudson M, Fritzler MJ, et al. Long-term exposure to a mixture of industrial SO2, NO2, and PM2.5and

anti-citrullinated protein antibody positivity. Environ Heal A Glob Access Sci Source. 2020;19(1):1–8.

319. Bernatsky S, Smargiassi A, Joseph L, Awadalla P, Colmegna I, Hudson M, et al. Industrial air emissions, and proximity to major industrial emitters, are associated with anti-citrullinated protein antibodies. Environ Res. 2017;157:60–3.
320. Favé MJ, Lamaze FC, Soave D, Hodgkinson A, Gauvin H, Bruat V, et al. Gene-by-environment interactions in urban populations modulate risk phenotypes. Nat Commun. 2018;9(1).
321. Rider CF, Yamamoto M, Günther OP, Hirota JA, Singh A, Tebbutt SJ, et al. Controlled diesel exhaust and allergen coexposure modulates microRNA and gene expression in humans: Effects on inflammatory lung markers. J Allergy Clin Immunol. 2016;138(6):1690–700.
322. Rider CF, Carlsten C. Air pollution and DNA methylation: Effects of exposure in humans. Clin Epigenetics. 2019;11(1):1–15.
323. Yun Y, Hou L, Sang N. SO2 inhalation modulates the expression of pro-inflammatory and pro-apoptotic genes in rat heart and lung. J Hazard Mater. 2011;185(1):482–8.
324. Thompson AMS, Zanobetti A, Silverman F, Schwartz J, Coull B, Urch B, et al. Baseline repeated measures from controlled human exposure studies: Associations between ambient air pollution exposure and the systemic inflammatory biomarkers IL-6 and fibrinogen. Environ Health Perspect. 2010;118(1):120–4.
325. Wooding DJ, Ryu MH, Li H, Alexis NE, Pena O, Carlsten C. Acute air pollution exposure alters neutrophils in never-smokers and at-risk humans. Eur Respir J. 2020;55(4).
326. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants with Those of the General Population. Am J Epidemiol. 2017;186(9):1026–34.
327. Giorgiutti S, Dieudonne Y, Hinschberger O, Nespola B, Campagne J, Rakotoarivelo HN, et al. Prevalence of Antineutrophil Cytoplasmic Antibody–Associated Vasculitis and Spatial Association With Quarries in a Region of Northeastern France: A Capture–Recapture and Geospatial Analysis. Arthritis

Rheumatol. 2021;73(11):2078–85.
328. Sano T, Makino N, Aoyama Y, Ae R, Kojo T, Kotani K, et al. Temporal and geographical clustering of Kawasaki disease in Japan: 2007–2012. Pediatr Int. 2016;58(11):1140–5.
329. Ballester J, Borràs S, Curcoll R, Navarro-Gallinad A, Pozdniakova S, Cañas L, et al. On the interpretation of the atmospheric mechanism transporting the environmental trigger of Kawasaki disease. PLoS One. 2019;14(12):1–7.
330. Burns JC, Cayan DR, Tong G, Bainto E V., Turner CL, Shike H, et al. Seasonality and temporal clustering of Kawasaki syndrome. Epidemiology. 2005;16(2):220–5.
331. Rypdal M, Rypdal V, Burney JA, Cayan D, Bainto E, Skochko S, et al. Clustering and climate associations of Kawasaki Disease in San Diego County suggest environmental triggers. Sci Rep. 2018;8(1):16140.
332. Information Services Division of Scotland. Community Health Index Number. 2023.
333. Zhang L, Yi H, Sang N. Sulfur dioxide-induced exacerbation of airway inflammation via reactive oxygen species production and the toll-like receptor 4/nuclear factor-κB pathway in asthmatic mice. Toxicol Ind Health. 2021;37(9):564–72.
334. Favé MJ, Lamaze FC, Soave D, Hodgkinson A, Gauvin H, Bruat V, et al. Gene-by-environment interactions in urban populations modulate risk phenotypes. Nat Commun. 2018;9(1).
335. Nanki T, Takada K, Komano Y, Morio T, Kanegane H, Nakajima A, et al. Chemokine receptor expression and functional effects of chemokines on B cells: Implication in the pathogenesis of rheumatoid arthritis. Arthritis Res Ther. 2009;11(5):1–11.
336. Ural BB, Caron DP, Dogra P, Wells SB, Szabo PA, Granot T, et al. Inhaled particulate accumulation with age impairs immune function and architecture in human lung lymph nodes. Nat Med. 2022;28(12):2622–32.
337. Cook CE, Fu X, Zhang Y, Stone JH, Choi HK, Wallace ZS. Validation of Antineutrophil Cytoplasmic Antibody–Associated Vasculitis as the Cause of End-Stage Renal Disease in the US Renal Data System. ACR Open Rheumatol. 2022;4(1):8–12.
338. Meineke EK, Daru BH. Bias assessments to expand research harnessing

biological collections. Trends Ecol Evol. 2021;36(12):1071–82.
339. Mahr A, Artigues N, Coste J, Aouba A, Pagnoux C, Guillevin L. Seasonal variations in onset of Wegener's granulomatosis: Increased in summer? [3]. J Rheumatol. 2006;33(4):1615 – 1622.
340. Barnett, Adrian; Dobson A. Analysing Seasonal Health Data. Brisbane: Springer Heidelberg Dordrecht London; 2009. 549 p.
341. Savige J, Gillis D, Benson E, Davies D, Esnault V, Falk RJ, et al. International consensus statement on testing and reporting of antineutrophil cytoplasmic antibodies (ANCA). Am J Clin Pathol. 1999;111(4):507–13.
342. Hagen EC, Daha MR, Hermans J, Andrassy K, Csernok E, Gaskin G, et al. Diagnostic value of standardized assays for anti-neutrophil cytoplasmic antibodies in idiopathic systemic vasculitis. Kidney Int. 1998;53(3):743–53.
343. Jennette JC, Falk RJ, Andrassy K, Bacon PA, Churg J, Gross WL, et al. Nomenclature of Systemic Vasculitides. Arthritis Rheum. 1994;37(2):187–92.
344. Nilsen AT, Karlsen C, Bakland G, Watts R, Luqmani R, Koldingsnes W. Increasing incidence and prevalence of ANCA-associated vasculitis in Northern Norway. Rheumatol (United Kingdom). 2020 Sep 1;59(9):2316–24.
345. Mohammad AJ, Segelmark M, Smith R, Englund M, Nilsson JÅ, Westman K, et al. Severe infection in antineutrophil cytoplasmic antibody-associated vasculitis. J Rheumatol. 2017;44(10):1468–75.
346. Y. O, I. U, K. M, T. K, K. S, K. M, et al. Incidence rate of anti-neutrophil cytoplasmic antibody (ANCA)-related vasculitis in pneumoconiosis and asbestosis patients in Japan-a retrospective study of 34 Rosai hospital group admission records. Am J Respir Crit Care Med. 2017;195:3872.
347. Injuries I, Council A. Occupational exposure to silica or asbestos and antineutrophil cytoplasmic antibodies ( ANCA ) - associated vasculitis : position paper 46. 2022;1–15.
348. Thygesen LC, Hvidtfeldt UA, Mikkelsen S, Brønnum-Hansen H. Quantification of the healthy worker effect : a nationwide cohort study among electricians in Denmark. BMC Public Health. 2011;11(57):1–11.
349. Aguirre-Gamboa R, Joosten I, Urbano PCM, van der Molen RG, van Rijssen E, van Cranenbroek B, et al. Differential Effects of Environmental and Genetic Factors on T and B Cell Immune Traits. Cell Rep. 2016;17(9):2474–87.
350. Madaniyazi L, Li S, Li S, Guo Y. Candidate gene expression in response to

low-level air pollution. Environ Int. 2020;140(January):105610.
351. Champion T. Focus on People and Migration. Focus People Migr. 2005;(January 2005).
352. Dennett A, Stillwell J. Population turnover and churn: enhancing understanding of internal migration in Britain through measures of stability. Popul Trends. 2008;(134):24–41.
353. Burns JC, Kushner HI, Bastian JF, Shike H, Shimizu C, Matsubara T, et al. Kawasaki disease: A brief history. Pediatrics. 2000 Aug;106(2):E27.
354. Banday AZ, Arul A, Vignesh P, Singh MP, Goyal K, Singh S. Kawasaki disease and influenza—new lessons from old associations. Clin Rheumatol. 2021;40(7):2991–9.
355. Makino N, Nakamura Y, Yashiro M, Sano T, Ae R, Kosami K, et al. Epidemiological observations of Kawasaki disease in Japan, 2013–2014. Pediatr Int. 2018;60(6):581–7.
356. Wallace ZS, Stone JH. Personalized Medicine in ANCA-Associated Vasculitis ANCA Specificity as the Guide? Front Immunol. 2019;10(December):1–10.
357. Public Health England. Estimation of costs to the NHS and social care due to the health impacts of air pollution: summary report About Public Health England. 2018;14 pp.
358. National Atmospheric Emissions Invetory. About Sulphur Dioxide [Internet]. NAEI. 2022 [cited 2023 Jun 15]. p. 1.
359. Juginović A, Vuković M, Aranza I, Biloš V. Health impacts of air pollution exposure from 1990 to 2019 in 43 European countries. Sci Rep. 2021;11(1):1–15.
360. Public Health England. Review of interventions to improve outdoor air quality and public health. PHE Publ. 2019;2–262.
361. Orellano P, Reynoso J, Quaranta N, Bardach A, Ciapponi A. Short-term exposure to particulate matter (PM10 and PM2.5), nitrogen dioxide (NO2), and ozone (O3) and all-cause and cause-specific mortality: Systematic review and meta-analysis. Environ Int. 2020;142(December 2019):105876.
362. European Commission. European Green Deal: Commission proposes rules for cleaner air and water. Europa.eu. 2022. p. 1.
363. Kelly F, Armstrong B, Atkinson R, Anderson HR, Barratt B, Beevers S, et al. The London low emission zone baseline study. Res Rep Health Eff Inst.

2011;(163):3–79.
364. Panteliadis P, Strak M, Hoek G, Weijers E, van der Zee S, Dijkema M. Implementation of a low emission zone and evaluation of effects on air quality by long-term monitoring. Atmos Environ. 2014;86(2):113–9.
365. Health Effect Institute, European Respiratory Society, International Society for Environmental Epidemiology. Brussels Meeting on Air Pollution and Health: Taking stock of the proposed revision to the Ambient Air Quality Directive [Internet]. HEI. 2023 [cited 2023 Jun 3]. p. 1.